Journal of
Experimental Psychology

VOL. 37, NO. 1 FEBRUARY, 1947

CONDITIONING AS A FUNCTION OF THE TIME BETWEEN CONDITIONED AND UNCONDITIONED STIMULI*

BY GREGORY A. KIMBLE†

State University of Iowa

INTRODUCTION

Theoretical background

One of the major corollaries of Hull's recent theoretical formulation (9) is concerned with the relation of the rate of conditioning to the time interval separating the onset of the conditioned stimulus (*Sc*) and the unconditioned stimulus (*Su*). It states: "In a reinforcement situation, there is a temporal relationship of the conditioned to the unconditioned stimulus such that as the onset of the unconditioned stimulus is progressively delayed, the rate of learning will increase" (p. 168). As Hull goes on to elaborate, this relationship holds only for short intervals, the maximum of the implied function being reached at about 500 msc. Beyond this point, a new or different relationship holds. The present study is concerned only with the first of these relationships.

Behind the above stated corollary, and apparently considered by Hull as the basis of its derivation, is a quasi-neurological hypothesis, which has been stated as follows: "Other things equal, the increment to the strength of a receptor-effector connection ($\Delta_S H_R$) resulting from a reinforcement is an increasing function of the frequency of the associated receptor discharge or the intensity of the resulting afferent impulse" (9, p. 168). This afferent discharge is, in turn, postulated to be a function, among other things, of the duration of the stimulus

* This paper constitutes the major portion of a dissertation submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy, in the Department of Psychology in the Graduate College of the State University of Iowa. The writer wishes to express his sincere appreciation to Professor Kenneth W. Spence for his many helpful suggestions and stimulating criticisms offered throughout the course of the experimental investigation.

†Brown University, Providence, R. I.

which produces it. This latter assumption is based upon experimental physiological studies which have shown that, when a receptor organ is stimulated, there result, after a brief period (latency), receptor or neural impulses, the frequency of discharge of which increases rapidly to a maximum. It is the frequency of these neural discharges which Hull considers a critical factor in determining the limit and rate of conditioning. Conditioning would be most efficient, according to this conception, when the *Sc-Su* interval is such that the receptor discharge from *Sc* is at its maximal frequency at the time of occurrence of the unconditioned response. According to Hull's hypothesis, this maximum is reached somewhere within one second.

A basic fault in this formulation becomes apparent, however, when attention is called to the fact that the physiological data cited by Hull as demonstration of the characteristics of the receptor neural discharge were data obtained from the reactions of cold-blooded animals, e.g., the eye of the eel. It is a well-established fact that the temporal characteristics of the neural activity of such specimens differ markedly from those of warm-blooded animals, being much slower in latency, longer in chronaxie, etc. Riggs (13), for example, has published electroretinographic records from the human eye which show the receptor discharge reaching a maximum at about 100 msc. Bartley (1) presents materials which indicate that the afferent neural discharge in the optic nerve of the rabbit has about the same temporal characteristics as the receptor discharge recorded electroretinographically. Both functions reach their maxima in approximately 100 msc. to 200 msc. after the onset of stimulation. On the assumption that conditioning is a function of the receptor discharge or the afferent neural discharge resulting from it, we should expect optimal conditioning to be reached at some point within this interval.[1] This is considerably short of the interval of 450 msc., which is most often cited as being optimal for conditioning. It is probably little more than accident that the afferent discharge pattern resulting from the stimulation of the peripheral organs of these animals is in correspondence with results which have characteristically been obtained in conditioning experiments with mammalian Ss.[2]

[1] If one assumes that maximum frequency of sensory discharge must coincide with the occurrence of the *response*, the optimal *Sc-Su* interval would be even smaller than the above figures indicate, for allowance would have to be made for the latency of the unconditioned response. If the latency were, for instance, 50 msc. the optimal interval would be reached somewhere between 50 msc. and 150 msc.

[2] It should not be concluded, however, that this unfortunate choice of neurological material entirely negates Hull's theoretical formulation. The mathematical part of the theory (which is the essential part of it) does not stand or fall with the proof or disproof of the neurophysiological notions which lie behind it. The critical tests of the theory must eventually come in the form of experimental studies designed to test the theoretical (mathematical) statements either directly or else in the form of tests of the deductions made on the basis of the theory.

Furthermore, there is certain indirect evidence which would suggest that the maximum rate of receptor discharge in the case of the human eye is attained considerably before the optimal interval for conditioning (300 msc. to 500 msc.). This evidence is to be found in the studies which have been concerned with the increase in apparent (reported) brightness of a visual stimulus as a function of the duration of stimulation. McDougall (11), for instance, found that for a *threshold* stimulus, maximum apparent brightness was obtained in approximately 200 msc., the apparent brightnesses of more intense stimuli reaching their maximum in even less time. A very extensive and carefully performed study by Bills (3) resulted in similar findings. The apparent brightnesses of visual stimulations reached their maxima in times ranging from less than 100 msc. to slightly more than 200 msc. depending upon the brilliance and hue of the light stimulus employed. It would seem reasonable to assume that the time taken for the afferent neural discharges from the human eye to reach a maximum would be no longer than that required for the perceptual response to reach its maximum.[3] On the basis of this assumption, at any rate, we would again be led to expect that maximum conditioning to a visual stimulus would occur at a much earlier interval than is at present considered to be the case.

Review of previous experimental studies on optimal Sc-Su interval

A number of studies have been reported which have attempted to establish the relationship between extent of conditioning and the time between conditioned and unconditioned stimuli. The present status of our knowledge about this relationship seems to suggest that the optimal time interval is somewhere in the neighborhood of 500 msc., and that either longer or shorter intervals are less effective in the establishment of conditioned responses. This generalization is based on the results of several studies.

The first is an investigation by H. M. Wolfle (14). She found that the maximum degree of conditioning of a finger withdrawal response was obtained when the time between *Sc* and *Su* was 500 msc. In this experiment, both longer and shorter intervals were less effective in the establishment of conditioned responses. Hull (9) has fitted a symmetrical bidirectional gradient to the Wolfle data and, by extrapolation, determined that, in this experiment, maximum conditioning would have been obtained at .44 sec. had such a time interval been used.

A second relevant study is that of Kappauf and Schlosberg (10).

[3] This assumption appears even more plausible when one considers that the perceptual response involves some central process of discrimination which would presumably consume a certain amount of time.

These investigators conditioned a complex response pattern in rats to the sound of a buzzer, by a delayed conditioning technique, the unconditioned stimulus being an electric shock. Depending upon which part of the response complex was chosen as the response measure, they found that most efficient conditioning occurred when the time between stimuli was either 667 msc. or 1000 msc. Although this study was not set up to answer very specifically the question as to the exact time interval at which most efficient conditioning is obtained, the results do suggest that the optimal interval is probably greater than $\frac{1}{3}$ sec. and less than one sec.

A third study is that of Reynolds (**12**). This investigation involved conditioning an air-puff elicited wink reflex to the sound of a click and varying the time between *Sc* and *Su*. The times employed in Reynolds' investigation were 250 msc., 450 msc., 1150 msc., and 2250 msc. Maximal conditioning was obtained at 450 msc. The 250 msc. interval, as well as the two intervals longer than 450 msc. resulted in a lesser degree of conditioning. Reynolds has supplied us with clear-cut evidence that, under favorable conditions, practically 100 percent conditioning is attainable when the *Sc-Su* interval is 450 msc. and that as the *Sc-Su* interval becomes progressively longer or shorter, the maximum level of conditioning becomes less and less.

Somewhat different results are to be found in a second study by Wolfle (**15**), who, in a situation essentially similar to that employed in her earlier study, ascertained the optimal time interval to be 200 mcs. Both longer and shorter intervals were found to be less effective. Still different results have been reported by Bernstein (**2**) who obtained most effective conditioning at 300 msc. for eyelid conditioning and found no significant differences between conditioning obtained at that interval and conditioning obtained at longer intervals up to 1480 msc. Intervals shorter than 300 msc. were found to be definitely inferior. Unfortunately, however, no groups were conditioned at intervals between 300 msc. and 500 msc., the exact region where the optimal interval is suspected to occur.

All of these studies, with the exception of that of Reynolds, have had two rather serious shortcomings which make their data difficult to interpret. The first of these concerns the manner of presentation of the data. In none of these studies, except Reynolds', are learning curves presented. Rather, the convention has been to present the percentage of conditioned responses obtained in the whole acquisition series.[4] Such being the case, it is impossible to obtain from the figures any indication of the maxima which the response curves approach during conditioning or of the rate at which the maxima are

[4] Kappauf and Schlosberg (**10**) present data in a slightly more satisfactory manner, in terms of number of conditioned responses obtained on the last three experimental days.

approached. It is quite possible for the total number of conditioned responses in a training series to be identical for a number of experimental groups, and, at the same time, for the asymptotes and rates of the learning curves to be markedly different.

The second difficulty with most of these previous studies has been that conditioning has taken place under extremely *massed* conditions of practice. The experimental technique commonly employed has been to present conditioning trials as rapidly as the apparatus will permit. The time between trials, for instance, in the Wolfle studies was 8 sec.–12 sec. In the Bernstein study, trials were given, on the average, about once every 15 sec.–20 sec. As Reynolds (12) has clearly demonstrated, the effect of massing of practice ordinarily is to reduce the level of conditioning possible by as much as 20 percent in the case of the conditions of massing and distribution used in his experiment. Such being the case, massing of practice will necessarily tend to obscure the effectiveness of conditioning, whatever the experimental conditions employed. This fact is particularly important when the experimental conditions are such as to lead to relatively *poor* conditioning, as in the case where the time between conditioned and unconditioned stimuli is of brief duration.

A further special problem that arises in connection with the study of the relation between *Sc-Su* time and the level of conditioning attainable involves the method by which the occurrence or non-occurrence of the conditioned response is specified. In the typical conditioning experiment, it has been customary to score as positive instances of conditioning all responses to *Sc* that anticipate the onset of *Su*. When the *Sc-Su* interval is not too short (e.g., 450 msc.) this measure of conditioning involves no serious difficulty, for as Hilgard (6) and Campbell (4) have shown, the upper limit of the distribution of latencies of eyelid CRs falls at about 400 msc. which is well within the *Sc-Su* interval. As this interval is decreased, however, it is obvious that fewer and fewer conditioned responses would be scored under this procedure, simply because the responses at the upper end of the latency distribution would not have an opportunity to anticipate *Su*. In other words, such a definition of a conditioned response results in the exclusion of a larger and larger proportion of potential conditioned responses as the *Sc-Su* interval is decreased. About the only method of avoiding this difficulty would appear to be the use of periodic test trials in which the unconditioned stimulus is omitted. This method was employed in the present experiment.

Purpose of the Experiment

The present investigation was directed at a more precise determination of the nature of the relationship holding between the *rate*

and *maximum level* of conditioning and the time interval between the conditioned and the unconditioned stimuli when the inter-stimulus interval is 400 msc. or less. The study was also directed at an investigation of the bearing of the experimental data to some of the hypothetical properties ascribed by Hull to the input segment of his concept of stimulus trace. To this end, the eyelid reflex of six different groups of subjects was conditioned to a light stimulus at six different *Sc-Su* intervals between 100 msc. and 400 msc.

Experimental Technique

Apparatus

The apparatus used was the Dodge pendulum-photochronograph which has been described elsewhere (5).[5] Eyelid movement was recorded by the paper eyelash method. The unconditioned stimulus was an air puff delivered by the fall of a 90 mm. column of mercury in a manometer. The air tube which delivered the puff to the *S*'s eye was divided so as to activate a plastic reed which recorded a shadow on the photographic record. Calibration of the moment of arrival of the puff at the *S*'s eye and the recorded movement of the reed showed no difference between the two within the limits of the recording system used.

The conditioned stimulus was a circular light, six cm. in diameter, reddish in color, and at a distance of 75 cm. from the eye of the *S*. Prior to its onset, the *S* was barely able to see the white circle which was produced by the milk-glass covering the circular opening of the light box. The apparent brightness of the conditioned stimulus, as measured by the MacBeth illuminometer, was .53 millilambert. The source of light was two GE neon lamps of 1½ watt intensity, chosen because they attained their maximum brilliance in less than 1/1000 sec. after the application of current to them. This feature seemed particularly desirable since the interval between stimuli in some cases was shorter than the time required for many standard filament lamps to reach maximum intensity. In order to keep *Sc* constant from group to group, some lamp such as the one used seemed necessary. In the conditioning situation, the conditioned stimulus was presented for 1500 msc. on each trial. The unconditioned stimulus was presented at the appropriate interval after the onset of the conditioned stimulus.

The selection of time intervals to be investigated was made in the following manner: it had been previously determined that the *Sc-Su* times to be used were to fall between 100 msc. and 400 msc. Accordingly, a portion of each of the 100 msc. and 400 msc. groups was run first. Upon the basis of the performances of these *S*s, it was decided that the 200 msc. and 300 msc. intervals should be used. When approximately one-half the total number of *S*s to be run in these groups had been completed, the desirability of running groups of *S*s at 225 msc. and 250 msc. became apparent.

Subjects

Sixty-nine *S*s in all were used in the experiment. They were divided into six groups, each group being conditioned at a different *Sc-Su* interval. Ten *S*s were assigned to each of the three longest interval groups, and 13 *S*s were assigned to each of the shortest interval groups. The assignment of *S*s to the various groups was controlled to the following extent: from four to seven *S*s were first run in each group, the exact number depending upon the number of *S*s available at that particular time. When these *S*s had been completed, enough *S*s were run in each group to bring the total number in each group to ten. Three additional *S*s were then run in each of the shortest interval groups. Complete randomization was not used because of the difficulties in setting and calibrating the apparatus that such a procedure would have entailed.

Sixty of the 69 *S*s were obtained in a class in the Psychology of Adjustment at the State University of Iowa. All had previously had a course in elementary psychology which had included a discussion of conditioning. The remaining nine *S*s were solicited in a course in ele-

[5] The apparatus, with such modifications as are indicated in the text, was that used by Reynolds (12).

mentary psychology prior to any classroom exposure to the principles of conditioning. These nine *Ss* were those used to bring the total number of *Ss* in the shortest interval groups up to 13. No important differences were observed in the performance of these *Ss* as compared with the more sophisticated *Ss*, except that many of the latter group guessed the purpose of the experiment. All *Ss* were undergraduate women.

Procedure

Each *S* was given 60 reinforced acquisition trials followed by 20 non-reinforced extinction trials. The number of acquisition trials was selected after an examination of Reynolds' (12) results had indicated that, for his *Ss*, conditioning seemed to have reached a maximum after approximately this number of trials under the conditions of distribution used in the present study. Both conditioning and extinction were carried out under conditions of relative distribution of practice, trials being separated by an interval of one to two min., with a mean inter-trial time of 1½ min. Each *S* served in two sessions given on successive days. The time between sessions varied from 20–28 hours, the average time being approximately 24 hours.

In both acquisition and extinction, photographic records were made only on every other trial (the 'even' trials) because of a lack of the necessary materials in sufficient quantity to make records on every trial. On every tenth acquisition trial the unconditioned stimulus was omitted. These trials were designated as 'test-trials.'

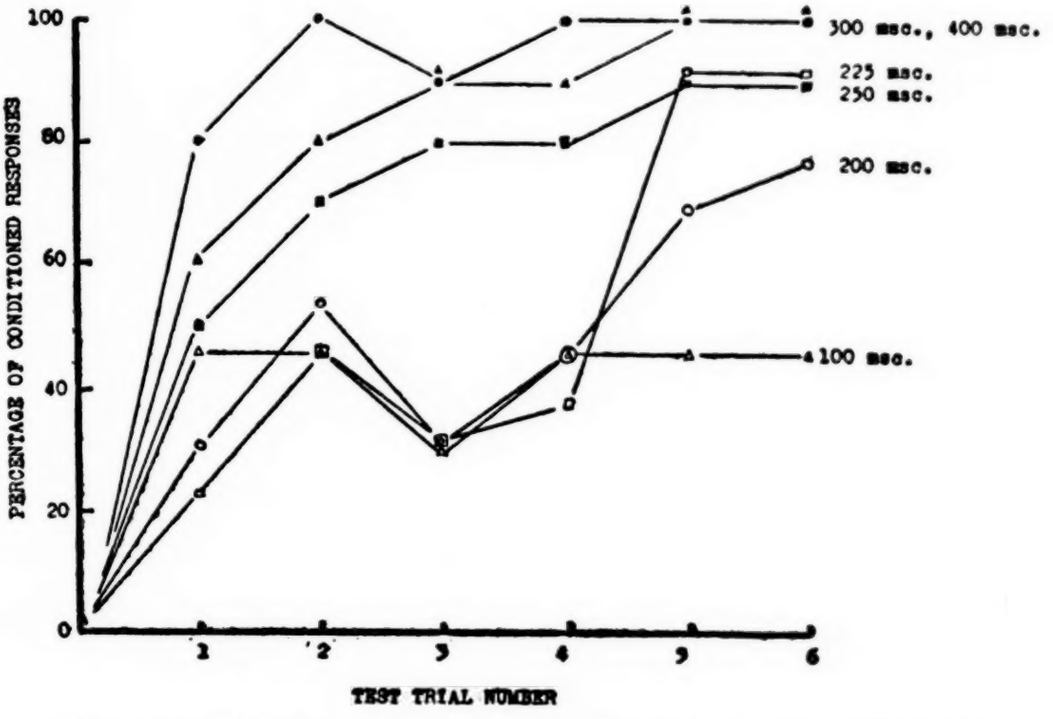

FIG. 1. Acquisition curves showing the percentage of *Ss* in each group giving conditioned responses on each test trial

RESULTS

Curves of acquisition

Acquisition curves based on the percentage of *Ss* showing a conditioned response at each of the test trials throughout the conditioning series are presented in Fig. 1. It will be observed that the curves, except in the case of those for the 200 msc. and 225 msc. groups, are fairly smooth and that they tend to be negatively accelerated in form. Also apparent is the fact that most of the curves seem to have reached

asymptotes, the doubtful instances being the irregular curves of the 200 and 225 msc. groups. The dips in these latter two curves at the third and fourth test trials do not appear to have any ready explanation other than being chance variations resulting from the small size of the samples.

Since many previous studies have employed measures of conditioning based on the concept of a conditioned response as an anticipatory response, curves of conditioning based on such a measure are presented in Fig. 2 so that they may be compared with similar data obtained in other studies. It will be noted from a comparison of Fig. 2 with Fig. 1 that, except for the 400 msc. group, all groups show

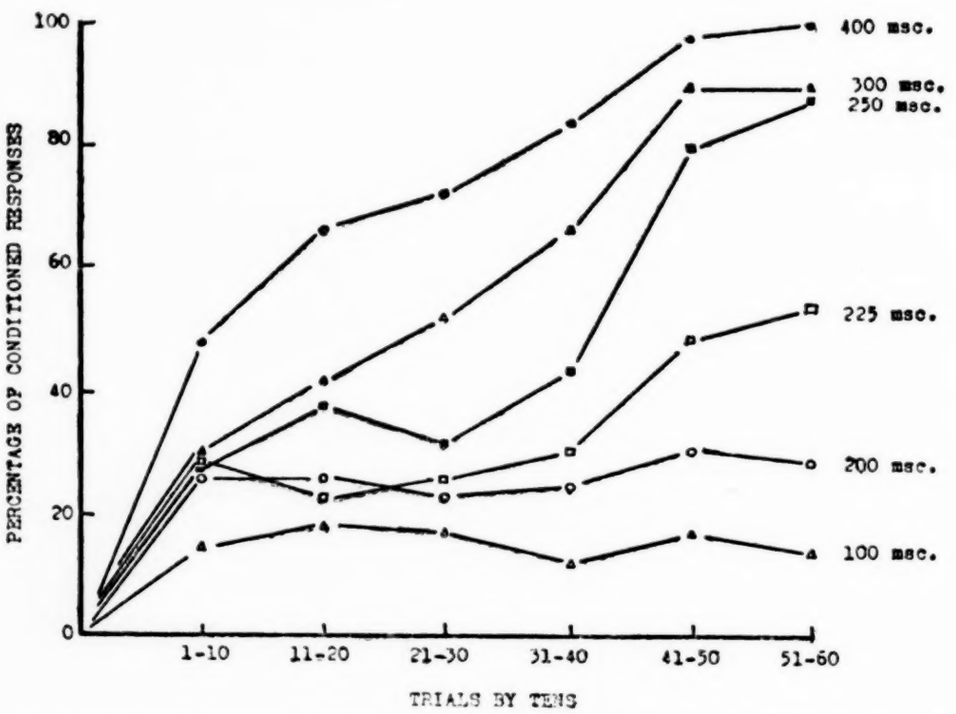

FIG. 2. Acquisition curves showing the mean percentage of anticipatory conditioned responses for each group in successive blocks of 10 trials. On test trials, only anticipatory responses were counted as conditioned responses

a lesser final degree of conditioning when the anticipatory response method of counting is used. Presumably this difference is a function of the fact that the anticipatory method of defining conditioned responses eliminates from the response measure all responses the latencies of which are greater than the *Sc-Su* interval.

Conditioning as a function of the Sc-Su interval

That both the limit and rate of acquisition of the conditioned response varied with the length of the *Sc-Su* time interval is readily apparent from the curves in Figs. 1 and 2. In both graphs the curve for the longest interval group tends to be consistently superior to the

others at each point, while the curve for the shortest interval group tends to show the poorest conditioning. Furthermore, while there is some crossing of the curves, there is an easily discernible tendency not only for them to end at different levels (asymptotes), but for this maximum value to be higher the longer the time interval.

A clearer picture of the nature of this latter relationship appears in Fig. 3. In this graph the level of conditioning attained by each group on the last test trial measured in terms of the percentage of *Ss* giving a conditioned response on that trial is plotted as a function of the *Sc-Su* interval. The broken line joining the experimental points reveals clearly that the maximum level of conditioning increased

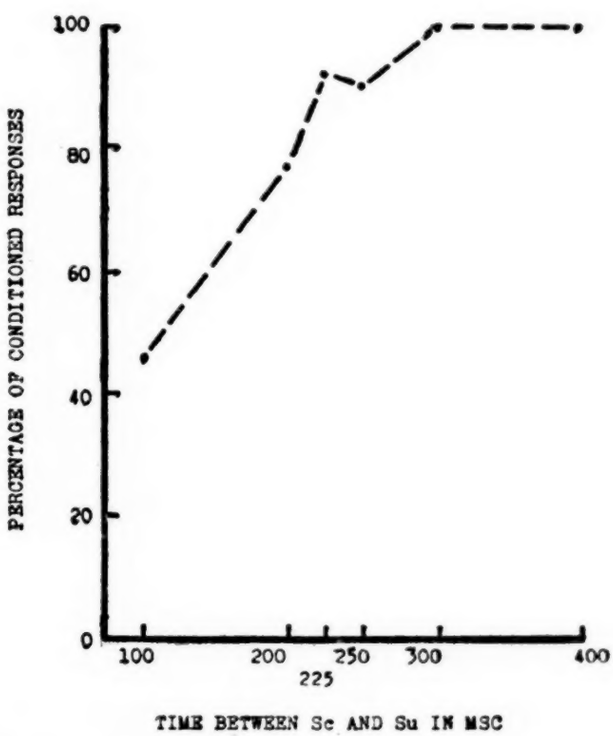

FIG. 3. Graph showing the limit of conditioning (percentage of *Ss* giving conditioned responses on the last test trial) as a function of the length of the *Sc-Su* interval

within the range of temporal intervals used as the length of the interval increased. Despite the small number of cases in each group, the curve is fairly regular, showing but a single inversion. It will be observed that the values for the two longest interval groups, 300 msc. and 400 msc., are the same. However, this may merely reflect the fact that a ceiling on the measuring scale has been reached. The two groups may have reached different levels of conditioning as measured by other indices.

That these two groups made a different total number of conditioned responses on all test trials is shown by the data of Table I. This table presents the mean percentage of conditioned responses

TABLE I

MEAN PERCENTAGE OF CONDITIONED RESPONSES ELICITED IN ALL TEST TRIALS

Group	Mean	S.D.
	Percentage of Conditioned Responses	
100	45	36.7
200	51	30.3
225	54	30.8
250	77	22.7
300	87	19.5
400	95	6.8

given by the Ss on all test trials, together with the standard deviations of these values. A *t*-test applied to the observed difference between the means of the 300 msc. and 400 msc. groups yields a *t*-value of 2.55 which is significant at the 2 percent level of confidence. This would tend to indicate that, according to this measure at least, there is a real difference in the amount of conditioning obtained in these two groups.

Further evidence on this point is provided by the data obtained during the extinction series. Since the unconditioned stimulus is omitted in extinction, every trial is a test trial and provides data. Table II shows the mean percentage of conditioned responses made

TABLE II

MEAN PERCENTAGE OF CONDITIONED RESPONSES ELICITED IN 20 EXTINCTION TRIALS

Group	Mean	S.D.
	Percentage of Conditioned Responses	
100	29	28.7
200	28	24.7
225	47	35.0
250	67	26.5
300	68	34.3
400	73	30.3

by the different groups in the 20 trial extinction series. As revealed by this measure, the strength of conditioning again appears not to have reached its maximum by 300 msc., but keeps on increasing up to the 400 msc. interval. The difference between the means obtained on this measure for the 300 msc. and 400 msc. groups is, however, not statistically significant, the *t*-value being .215.

Turning now to the matter of *rate* of conditioning, or the speed with which the conditioned response develops, Fig. 4 presents a picture of how one measure of this variable changes with the temporal period between the conditioned and unconditioned stimulus. As an index of rate of conditioning, the number of trials required for each group to attain 75 percent conditioning was noted. The curve in Fig. 4 was constructed by plotting 100 minus the number of trials required for each of the acquisition curves shown in Fig. 1 to reach this level.

This method of presentation enables us to represent the rate of conditioning as an *increasing* function of the *Sc-Su* interval without altering the relative values of the points plotted. No point is plotted for the 100 msc. group because 75 percent conditioning was not reached by this group. It will be observed that this curve is not dissimilar in shape to that presented in Fig. 3 where the limit of conditioning was plotted as a function of the *Sc-Su* interval. In

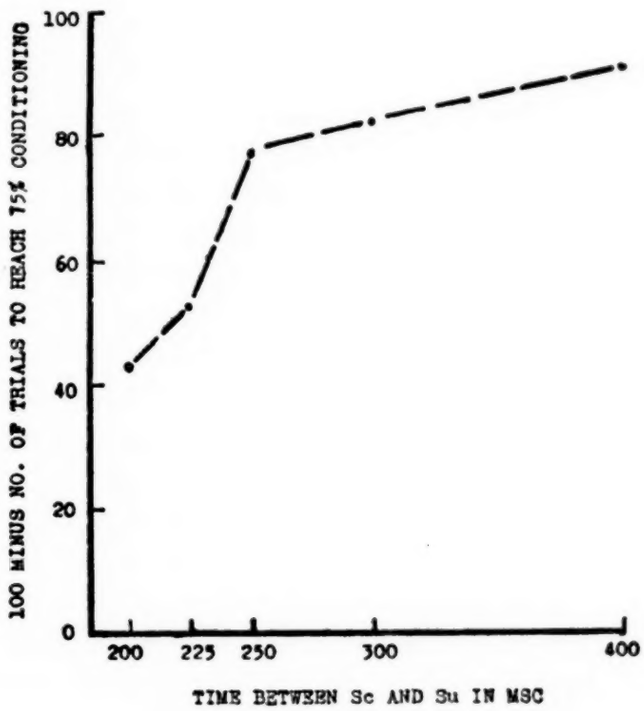

FIG. 4. Graph showing the rate of conditioning as a function of the length of the *Sc-Su* interval

contrast to Fig. 3, however, the rate of conditioning is seen to be greater for the 400 msc. group than for the 300 msc. group, whereas the limits of the two curves did not differ.

DISCUSSION OF RESULTS

In Hull's theoretical formulation, the time between *Sc* and the response is postulated as one of the many factors determining the asymptote of the habit growth curve.[6] The assumption made is that, when the time between the conditioned stimulus and the response is

[6] The magnitude of the response measure, e.g., percentage of responses, resistance to extinction, etc., theoretically depends upon a number of other variables besides the strength of the habit (*H*). These other variables include the motivation of the *S* (*D*), the total amount of inhibition developed during learning (I_R), and chance oscillation factors. Experimental controls were instituted, however, for the purpose of equating these factors from group to group. To the extent that this was actually accomplished, the response measure directly reflects the strength of the habit, and differences in response measures may be interpreted as reflecting differences in habit strength.

less than about 500 msc., the limit of rise of the habit curve is an increasing function of the length of this interval.

In general, the present findings confirm this assumption. The fact that different limits of conditioning are attained by the different groups definitely supports Hull's guess that this variable of the *Sc-Su* time interval determines the limit to which the hypothetical habit (*H*) curves grow. Reference to Fig. 3, however, suggests the possibility that the maximum of this function may be attained before 500 msc., perhaps as early as 300 msc. But it should be noted that our data on this latter point are not decisive. The fact that the 300 msc. and 400 msc. groups had the same maximum value, i.e., 100 percent, does not permit one to conclude that the limits of their hypothetical habit growth curves would necessarily be identical. Unfortunately, in the case of these two groups, we have come to the end of our measuring scale (percentage of response) and hence it is not possible to ascertain by means of this measure any evidence as to possible differences in the limits of their strengths. The more rapid learning exhibited by the 400 msc. group as revealed by the greater number of conditioned responses made in the whole series, and the smaller number of trials required to reach the 75 percent level of conditioning, coupled with the evidence of different limits of learning for the shorter interval groups, would lead one to infer that the *limit* of the habit growth curve would be higher for the 400 msc. group.

In the context of Hull's theoretical treatment of the habit growth function, there is one alternative to the hypothesis that it is the *limit* of the habit curve which is determined by the inter stimulus time variable. That alternative would be that the parameter determining the rate of approach of the habit curve to its maximum is greater in the case of the 400 msc. group than in the 300 msc. group. In an earlier formulation of the influence of this variable, Hull (**8**) postulated that the effect of its variation was to change this parameter of the exponential habit growth function and not the maximum. The weight of evidence, however, would appear to be in favor of the more recent assumption that it is the maximum that varies. Casual inspection of the curves of conditioning in Fig. 1 certainly does not reveal any tendency for the *rate of approach to the different maxima* (growth constants) to vary in any systematic fashion.[7]

A major remaining question concerns the specific nature of the function according to which the limit of habit growth is determined

[7] The reader should distinguish carefully between the *rate* measure that was defined in the preceding section in terms of the number of trials required to reach a specified level on the response measuring scale and the parameter discussed here which is a growth constant in the mathematical expression for the habit curve. Functions that are identical in this latter characteristic differ in the former if the two functions have different asymptotes.

by the length of the *Sc-Su* interval. Reference to Figs. 3 and 4 indicates that, for the present study, this function was found to be a negatively accelerated one which reaches its maximum at some point beyond 400 msc. The shape of the function is somewhat different from that fitted by Hull (9) to the data of one of the Wolfle (12) studies in his discussion of this point. The curve presented by Hull is one such that, as the time interval between *Sc* and *Su* becomes progressively longer or shorter than the optimal interval of 440 msc., the amount of conditioning decreases as a simple negative exponential function from this point. In terms of increasing time intervals, this means that the amount of conditioning is a positively accelerated function of the *Sc-Su* time for the portion of the curve between zero msc. and 440 msc. Although the Wolfle data were obtained in a trace conditioning experiment, whereas the present study involved a delay conditioning technique, there is, at present, no evidence to indicate that there should be any essential difference in the shape of this function for the two different types of conditioning experiment. In general, the available evidence suggests that the segment of the curve between simultaneity and the optimal interval exhibits negative acceleration as the maximum is approached. For example, Hilgard and Marquis (7, pp. 160–165) have presented temporal gradients for four relevant studies. An examination of these curves shows that every one of them is negatively accelerated during the portion of the curve under discussion, including the Wolfle study, the results of which Hull fitted with a positively accelerated gradient.

The point has previously been made that, in Hull's theoretical structure, the time between *Sc* and *Su* is one of the factors determining the limit of rise of the habit growth curve. In an earlier, unpublished formulation (7) of the rôle of the *Sc-Su* time interval in conditioning, Hull postulated that this interval was one of the variables which determine the magnitude of an hypothetical intervening variable which he called the *input segment of the stimulus trace* ($\dot{s}$). The value of this hypothetical variable was in turn assumed to determine the rate and degree of development of habit strength. Hull's complete definition of this theoretical construct was as follows:

$$\dot{s} = a \log S(1 - e^{-hT}) \qquad (1)$$

where $\dot{s}$ = the input segment of the stimulus trace, S = the intensity of the stimulus, T = the duration of S since its beginning, a and h = empirical constants.

If one assumes with Hull that there is a one-to-one relationship between the maximum amount of conditioning obtainable under different interstimulus time intervals and the value of this hypothetical variable, $\dot{s}$, it will be seen that the curves of Figs. 3 and 4

closely approximate the negatively accelerated exponential function represented by equation 1.

Finally, it may be noted that, while the results of the present investigation are in essential agreement with Hull's original mathematical statement of the relationship between stimulus duration and the magnitude of the stimulus trace, they do not confirm his added neurophysiological theorizing about this construct. Instead, they bear out the findings of most previous studies in suggesting that the optimal interval for conditioning is 400 msc. or longer, a value much greater than the time required for either the receptor or neural discharge to reach a maximum in the case of mammalian organisms. The fact that the optimal interval for conditioning is almost identical with the recruitment time of such functions in cold-blooded animals is probably no more than a chance similarity.

Summary

1. Six groups of human Ss were conditioned in an eyelid conditioning situation at six different *Sc-Su* intervals of 400 msc. and less, with the end in view of investigating the relationship between the level and rate of conditioning and the time between *Sc* and *Su*.

2. The results indicate that, for the range of intervals used in this experiment, the level of conditioning is an increasing, negatively accelerated function of the length of the *Sc-Su* interval.

3. The rate of conditioning is also an increasing, negatively accelerated function of the length of the *Sc-Su* interval.

4. The simplest interpretation of the results suggests that the length of the *Sc-Su* interval is one of the factors which determines the asymptote of the habit-growth curve.

5. These results and other evidence throw doubt upon the value of Hull's neurophysiological interpretation of his construct, stimulus trace, but do support the purely mathematical definition of it.

(Manuscript received February 19, 1946)

References

1. Bartley, S. H. *Vision.* New York: D. Van Nostrand Company, Inc., 1941.
2. Bernstein, A. L. Temporal factors in the formation of conditioned eyelid reactions in human subjects. *J. gen. Psychol.*, 1934, **10**, 173–197.
3. Bills, M. A. The lag of visual sensation in its relation to wave length and intensity of light. *Psychol. Monogr.*, 1920, **28**, No. 5, 101 pp.
4. Campbell, A. A. The interrelations of two measures of conditioning in man. *J. exp. Psychol.*, 1938, **22**, 225–243.
5. Dodge, R. A pendulum-photochronograph. *J. exp. Psychol.*, 1926, **9**, 155–161.
6. Hilgard, E. R. Conditioned eyelid reactions to a light stimulus based on the reflex wink to sound. *Psychol. Monogr.*, 1931, **41**, No. 184, 50 pp.
7. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: D. Appleton Century Company, 1940.

8. HULL, C. L. Psychology seminar memoranda. 1930–1940. Bound manuscripts on file at the Oberlin College library, the University of Iowa library, and the Yale University Library.
9. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century Company, 1943.
10. KAPPAUF, W. E., & SCHLOSBERG, H. Conditioned responses in the white rat. III. Conditioning as a function of the length of the period of delay. *J. genet. Psychol.*, 1937, **50**, 27–45.
11. MCDOUGALL, W. The variation of the intensity of visual sensations with the duration of the stimulus. *Brit. J. Psychol.*, 1904, **1**, 151–189.
12. REYNOLDS, B. The acquisition of a trace conditioned response as a function of the magnitude of the stimulus trace. *J. exp. Psychol.*, 1945, **35**, 15–30.
13. RIGGS, L. A. Continuous and reproducible records of the electrical activity of the human retina. *Proc. Soc. exp. Biol. Med.*, 1941, **48**, 204–207.
14. WOLFLE, H. M. Time factors in conditioning finger-withdrawal. *J. gen. Psychol.*, 1930, **4**, 372–378.
15. WOLFLE, H. M. Conditioning as a function of the interval between the conditioned and the original stimulus. *J. gen. Psychol.*, 1932, **7**, 80–103.

THE RELATION BETWEEN DISTRIBUTION OF PRACTICE AND LEARNING EFFICIENCY IN PSYCHOMOTOR PERFORMANCE

BY JOSEPH C. FRANKLIN* AND JOSEF BROZEK

The Laboratory of Physiological Hygiene, University of Minnesota

Introduction

The relative effectiveness of spaced versus concentrated practice is an important problem in the study of the learning process. This problem exists throughout a wide range of activities, from academic training to athletics. In industry the question arises when a worker is to be trained for an operation which is new to him or when he is learning a new method of performing a given task. Will the operation be most quickly mastered in concentrated periods of practice or in spaced practice sessions? This question leads to others. What effect does activity during intervals between practice sessions have upon the progress of training? Will a training program providing instructional talks or films between practice trials result in a greater improvement than continuous practice? What are the optimal intervals between practice periods? Attempts to improve learning efficiency in any field of activity must consider the problem of optimal spacing of training sessions.

This problem has been studied on a wide variety of materials: digits, nonsense syllables, vocabulary, code, artificial language, prose, poetry, music, geography, history, mirror-reading, color discrimination, mazes, puzzle boxes, pursuit meter, mirror-tracing, archery, and typewriting. For the most part the experimental studies have demonstrated greater efficiency of spaced as compared with massed practice (1, 2). As early as 1897, Jost (3), using nonsense syllables for learning material, obtained recall scores of 7, 31, and 55 when the 24 learning sessions were distributed over 3, 6, and 12 days, respectively. However, occasional contradictory experimental findings (4, 5, 6, 7, 8, 9, 10, 11) favoring concentration of practice illustrate the complexity of the problem and caution against making a general assumption that spaced practice is more efficient in all learning situations.

In studies of human fitness conducted at the Laboratory of Physiological Hygiene, University of Minnesota, a variety of 'motor'

* Now at Department of Psychology, Purdue University.

tests are used. In general, steady performance on these tests must be obtained in the pre-experimental period of standardization. Subsequent deviations in performance may then be interpreted in the light of experimental variables, such as caloric deficit. It is desirable to develop methods and techniques securing the greatest possible economy of training in order to reduce the demands upon subjects, staff, laboratory facilities and equipment. An efficient spacing of practice periods should decrease the number of practice sessions needed for reaching a stable plateau and thus minimize necessary cost, time, and effort.

This investigation was undertaken in order to study the influence of varying the intervals between practice trials in training subjects on two psychomotor tests.

Conditions of the Experiment

Subjects

Thirty-six 'normal' young men, ranging in age from 20 to 33 with a mean age of 25.7 years, served as *S*s. For purposes of a larger experimental program the men were maintained on a controlled dietary regimen. The diet was 'good' according to accepted standards and provided an average of 3150 calories a day. The *S*s lived in a dormitory within the Laboratory and according to rules eliminating any wide variation in their mode of life. Since the men were conscientious objectors who volunteered for participation in the larger experiment, the aim of which was to provide scientific information for furthering nutritional relief and rehabilitation, their motivation was probably superior to that of most *S*s routinely employed in experimental work. None of the *S*s was previously acquainted with the tests or had practiced the particular skills involved in the performance of the tests.

Methods

From the tests intended for use in the larger nutritional experiment and which were to be practiced to a plateau, the test of gross body reaction time and the pattern-tracing test of eye-hand-foot co-ordination were selected for the study of concentrated versus distributed learning.

Gross body reaction time was measured while the *S* walked on a treadmill set at an angle of 10 percent climb and moving at 3½ miles per hour. Five feet ahead of the *S* at about head-height three signal lights were mounted on the frame of the treadmill. When the lights were turned on green was at the left, white in the center, and red at the right. At each side of the *S* was a paddle-key mounted on a post 20 in. from the floor of the treadmill. A blind was placed 1½ feet in front of the three signal lights and was adjusted so that the *S* had always to assume an upright position in order to see the signal and then bend in order to tap the appropirate paddle-keys. The green light was turned off by tapping the left, the red by tapping the right, and the white by tapping both paddle-keys simultaneously. The light stimuli were presented in random order. Each stimulus followed the preceding response at intervals varying from one-half sec. to two sec. A chronometer registering hundredths of a sec. recorded the time elapsing between the presentation of each stimulus and the break made in the circuit by the *S*s depressing the proper paddle-keys. The reaction times were summated for each set of 25 reactions.

In performing the pattern-tracing test, the *S* was required to trace a grooved pattern using a stylus tipped with a round metal head. Whenever the tip of the stylus came into contact with the sides of the pattern, a contact was electrically recorded on a counter and the length of the contact in hundredths of a sec. registered on the chronometer. Contact with the side of the pattern was indicated to the *S* by the sounding of a buzzer. The length of the path was 73½ in. and the depth and width were $\frac{5}{16}$ in. In order to eliminate variations in performance which would be due to differences in the speed of tracing the pattern, all *S*s were trained to trace the path twice in 45 sec. 'Half-time' was called by the tester at 22½ sec. With practice, the *S*s' speed of tracing became fairly constant.

Each training session consisted of three 'rounds.' Each round was made up of 25 reactions on the gross body reaction-time and one administration of the pattern-tracing test. The tests were given in the sequence: reaction time, pattern tracing. The sequence was repeated three times, which required a total of about seven to eight min. Immediately preceding each training session the *Ss* 'warmed up' by walking for 10 min. on the treadmill at the same speed and angle as was used during the practice trials.

Design

The *Ss* were placed in six groups of six individuals each on the basis of their performance on a 'try-out' trial. The distribution of high through middle and low scorers was similar from group to group and the mean performance for all groups was substantially the same. Perfect matching was not possible since three scores were involved. Table I presents the 'try-out' mean scores of the six groups and the practice schedules.

TABLE I

GROUPS, SCORES ON PRELIMINARY TRIAL, TRAINING REGIMEN

Practice Schedule	Total Training Time	Reaction Time, in Sec.		Pattern Tracing			
				No. of Contacts		Length of Contact, in Sec.	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
A. Three times a day.....	6 days	.58	.045	124	5.6	13.2	1.34
B. Twice a day.....	9 days	.57	.058	132	8.2	12.9	1.95
C. Once a day.....	18 days	.57	.045	126	8.5	13.3	1.45
D. Three times a week.....	6 weeks	.57	.057	131	8.8	13.1	1.05
E. Irregularly.....	6 weeks	.58	.045	136	7.2	14.1	1.58
F. Irregularly.....	6 weeks	.57	.044	137	8.8	13.9	1.48

Intervals between practice trials were equally distributed in groups A through D. For the individuals in groups E and F the intervals between practice periods ranged from less than an hour to two weeks. No *Ss* were practiced over weekends.

RESULTS

In the reaction-time test the six groups have learning curves of very similar shape (Fig. 1). This indicates that no special advantage in learning efficiency attaches to any one form of distribution of practice periods. Analysis of variance (see Table II) substantiates inspectional interpretation. While all groups showed significant performance improvement in successive trials, there were no significant differences between the groups within trials.

Learning curves for the number of contacts in the pattern-tracing test are given in Fig. 2. The results of analysis of variance reveal that the apparent divergence between groups within trials is not significant.

The learning curves for length of contact in pattern tracing are shown in Fig. 3. Again, while all groups improved significantly in performance over the trials, the groups are not distinguished from each other within the practice periods.

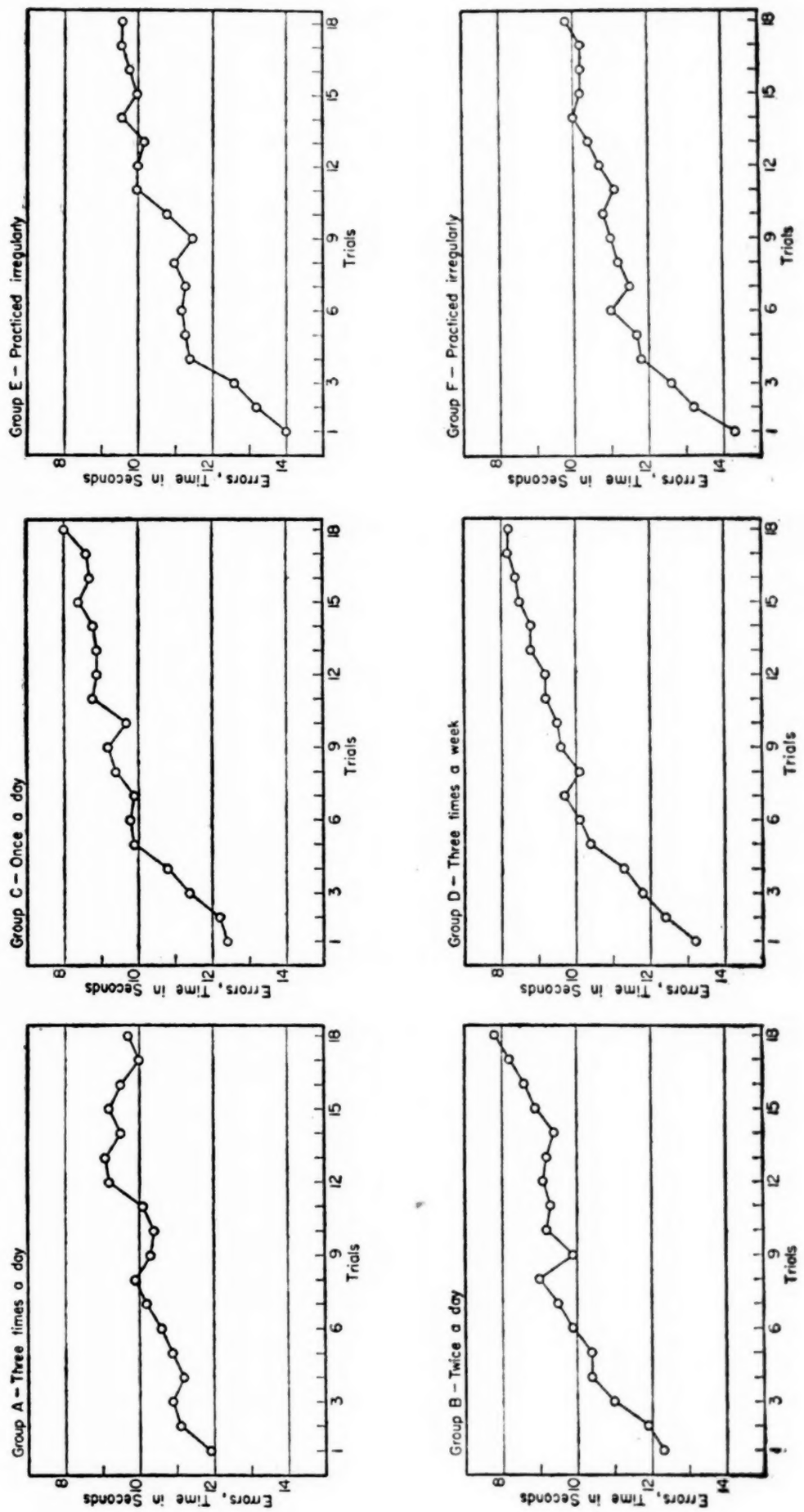

FIG. 1. Learning curves for the reaction-time test

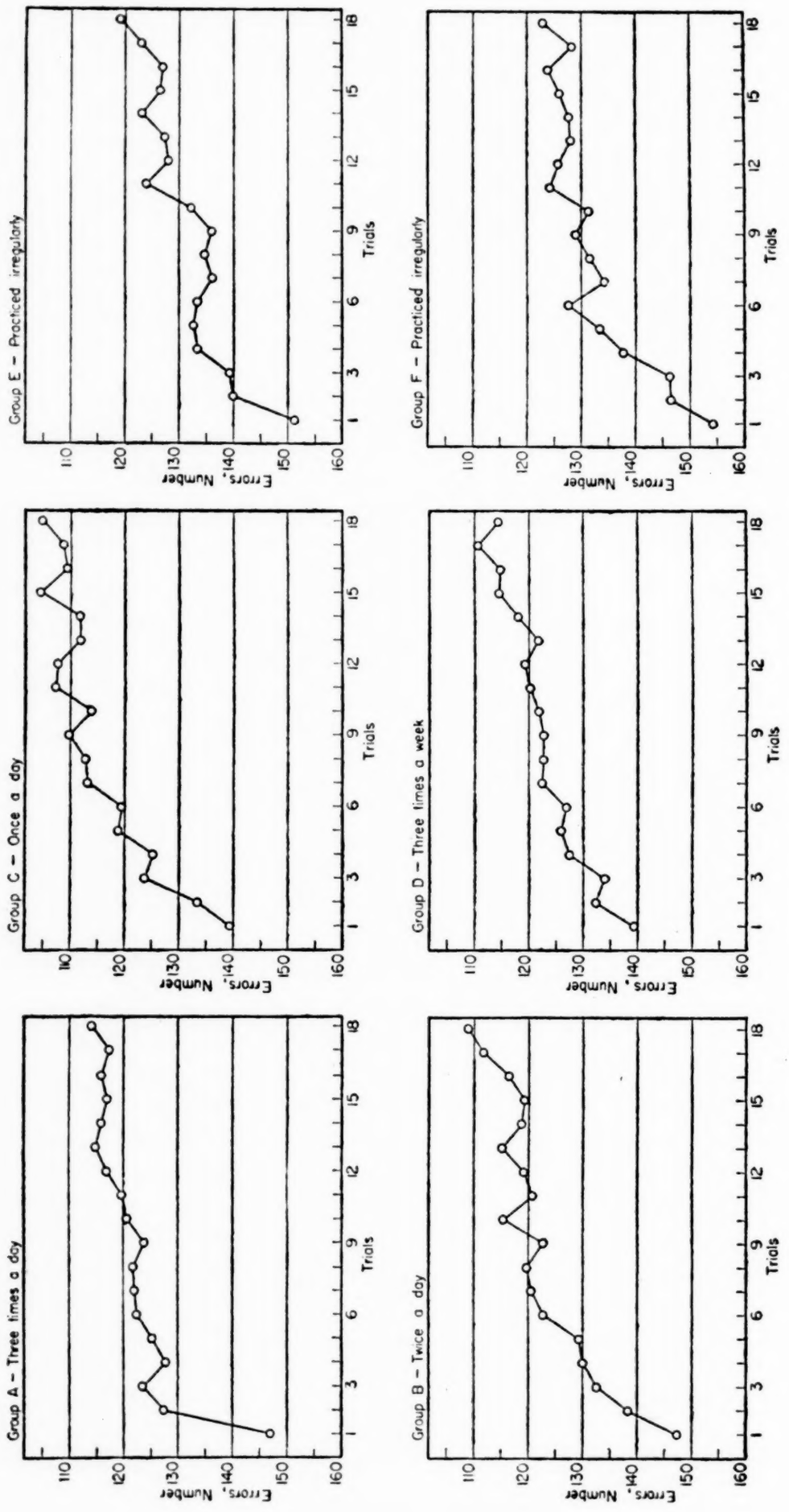

FIG. 2 Learning curves for pattern-tracing test, number of error contacts

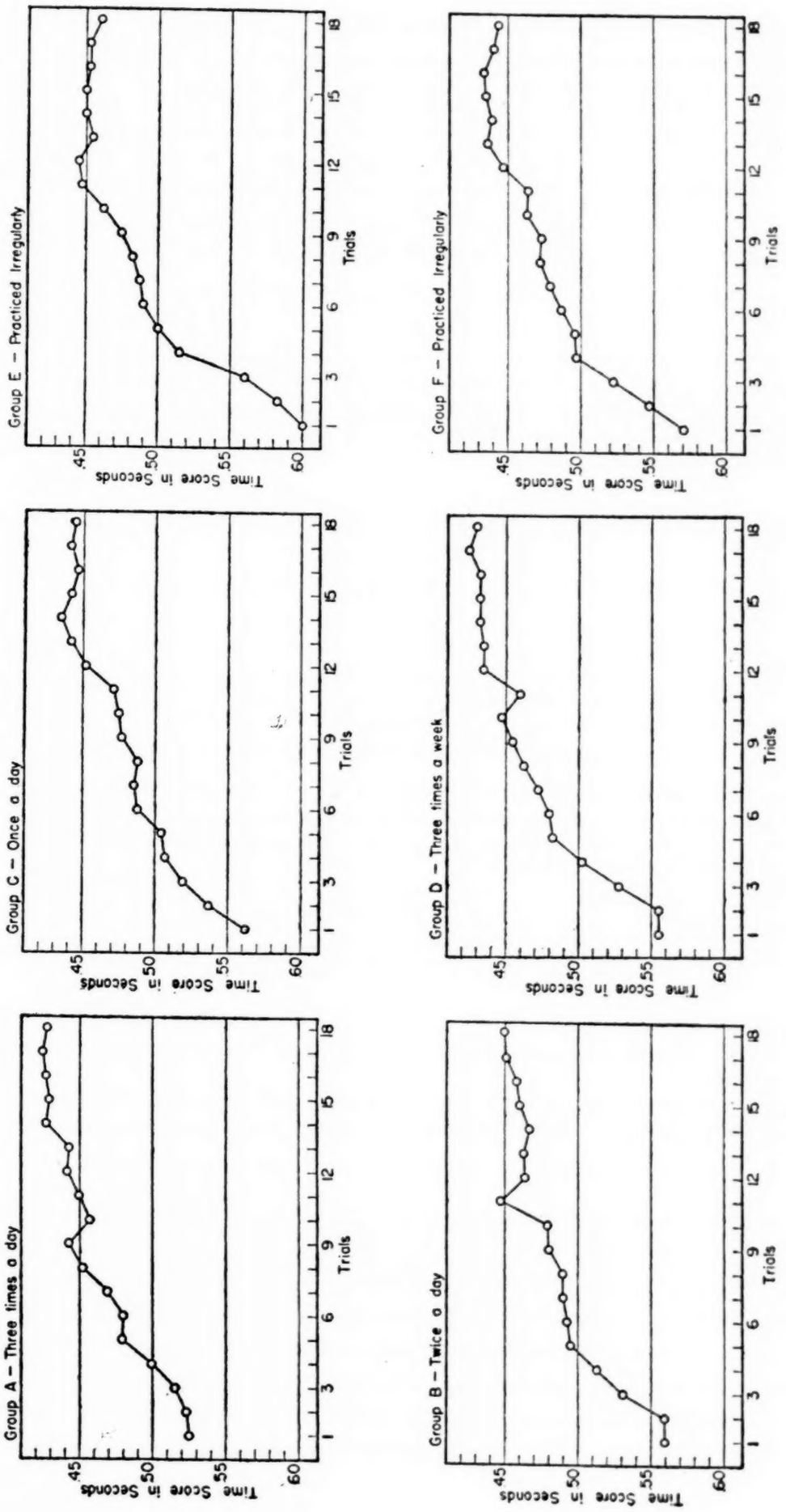

FIG. 3. Learning curves for pattern-tracing test, total time of error centacts

TABLE II

SIGNIFICANCE OF THE DIFFERENCES BETWEEN GROUP MEANS AT SPECIFIED TRIALS (1, 5, 9, 13, 18) AND OF THE DIFFERENCES BETWEEN THE GRAND MEANS OF THE FIRST AND THE LAST PRACTICE TRIAL

Function	F_{bG} for Trial					F_{bT} for Trials
	1	5	9	13	18	1 and 18
Reaction time	1.52	0.36	1.23	0.57	0.93	287.99*
Pattern tracing, number of contacts	1.25	1.47	2.28	1.33	1.42	194.69*
Pattern tracing, length of contacts	2.43	0.99	1.19	0.75	1.81	193.90*

* Significance of the differences *between* the means of the six *groups* at a given practice trial has been obtained in the form of a F-test, representing the ratio of the 'between groups' variance (V_{bG}) to the 'within groups' variance (V_{wG}), $F_{bG} = V_{bG}/V_{wG}$. The working formulae for computing V_{bG} and V_{wG} are as follows:

$$V_{bG} = \frac{\sum_{1}^{k}(\sum_{1}^{n} X_G)^2/n - (\sum_{1}^{nk} X)^2/n.k}{k-1}$$

$$V_{wG} = \frac{\sum_{1}^{nk} X^2 - \sum_{1}^{k}(\sum_{1}^{n} X_G)^2/n}{n(k-1)}$$

where
k = number of groups
n = number of individuals in each group
X = any individual score
X_G = individual score in a given group

The value of F associated with $(k-1) = 5$ and $n(k-1) = 30$ degrees of freedom is 2.53 at the 5% level of significance and 3.70 at the 1% level of significance.

Significance of the difference *between* the general mean of the initial and terminal *trials* is again tested by an F ratio. The formula is:

$$F_{bT} = \frac{(\sum_{1}^{nk} d)^2/nk}{\frac{\sum_{1}^{nk} d^2 - (\sum_{1}^{nk} d)^2/nk}{nk-1}}$$

where $d = X_{18} - X_1$, the difference between the individual's score in the eighteenth and the first practice trial.

For an F associated with 1 and $(nk-1) = 35$ degrees of freedom, the 5% point is 4.12, the 1% point is 7.42. An F value which reaches or exceeds these levels is marked with * and **, respectively.

Thus, the distribution of practice trials on the tests of gross body reaction time and pattern tracing varying from relatively massed practice—three trials a day—to irregular practice periods extended over six weeks disclosed no differences in learning efficiency.

On each of the three psychomotor tests the average scores of the six groups were not significantly different from each other within trials. This means, also, that the learning curves plotted from the trial averages of each of the six groups for any of the three psycho-

motor tests do not differ significantly and should be considered as random samples drawn from a homogeneous population of learning curves. By inference, the group differences in reaching the plateau in the gross body reaction time cannot be considered significant. In the other two tests, in which a true plateau had not been reached, a similarity in the time of reaching the plateau and in the level of plateau performance can be reasonably postulated if the training were carried further. We know from other experimental work that the performance in these two tests was actually extremely close to a true plateau.

Comment

At the present stage of knowledge the relative efficiency, if any, of massed or spaced practice in 'motor' learning cannot be predicted on the basis of theoretical considerations and must be determined by empirical means. Generalization from the results of previous studies of massed and spaced learning is severely limited by the difficulty of identifying the static and dynamic elements present in a given learning situation.

In this study the differential effects of massed practice as contrasted with those of distributed practice could have been obtained, if at all, only by scheduling the training periods somewhere between continuous practice and the most concentrated practice used. The massing of practice periods in these tests, however, is feasible only to a limited degree since the onset of tiredness in continued work is rapid and leads to deterioration of performance rather than to improvement through learning (**12**). Moreover, since other responsibilities of the Ss and the staff as well as the time limitations on the use of equipment precluded the administration of more than three trials a day to any S, the study of greater massing of practice trials must be left for the future.

From the results of the experiment it may be concluded:

1. There is no difference in efficiency of learning on these tests between relatively massed and distributed practice as measured by total number of practice trials needed to reach plateau.
2. Learning on these tests is as effective when trials irregularly scheduled are used as when sessions are strictly regular.

Since concentrated practice (three times a day) is as efficient as practice more widely spaced, intensive training may be undertaken without loss of learning effectiveness whenever it is desirable to train Ss in a minimum of time. The application of the finding that no loss of learning efficiency is occasioned by irregular scheduling of

practice sessions makes possible greater flexibility in the laboratory program since practice trials may be run at the convenience of *Ss* and staff.

Summary

The purpose of this experiment was to ascertain the relative learning efficiency of concentrated and distributed practice in training men on two psychomotor tests—gross body reaction time and pattern tracing.

Thirty-six young men served as *Ss*. They were matched on the basis of a 'try-out' performance on the tests. Six groups of six *Ss* each were practiced as follows: three times a day, twice a day, once a day, and three times a week; two groups were practiced irregularly.

Statistical analysis of the learning curves for all three measures of performance showed that no advantage in economy of learning was attached to any one of the schedules used. The number of trials required to achieve plateau performance was the same irrespective of the distribution of the practice periods.

These results can be applied to the practical problem of training *Ss* on these psychomotor tests. No loss of learning efficiency in training occurs in practicing *Ss* either intensively or extensively. Further, practice sessions may be scheduled irregularly with little regard for length of intervals between trials without impairment of learning effectiveness.

(Manuscript received January 18, 1946)

References

1. Ruch, T. C. Factors influencing the relative economy of massed and distributed learning. *Psychol. Rev.*, 1928, 35, 19–45.
2. McGeoch, J. A. *The psychology of human learning* (New York: Longmans, Green and Co., 1942), Chapt. IV, 119–154.
3. Jost, A. Assoziationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen. *Z. Psychol.*, 1897, 14, 436–472.
4. Carr, H. A. Distribution of effort. *Psychol. Bull.*, 1919, 16, 26–28.
5. Cook, S. A. The effects of various temporal arrangements of practice on the mastery of an animal maze of moderate complexity. *Arch. Psychol.*, 1928, No. 98.
6. Cook, T. W. Massed and distributed practice in puzzle-solving. *Psychol. Rev.*, 1934, 41, 330–355.
7. Cook, T. W. Distribution of practice and size of maze pattern. *Brit. J. Psychol.*, 1937, 27, 303–312.
8. Cook, T. W. Factors in massed and distributed practice. *J. exp. Psychol.*, 1944, 34, 325–334.
9. Cummins, R. A. Improvement and the distribution of practice. *Teach. Coll. Contr. Educ.*, 1919, No. 97.
10. Pechstein, L. A. Massed versus distributed effort in learning. *J. educ. Psychol.*, 1921, 12, 92–97.
11. Wingfield, R. C., & Dennis, W. J. The dependence of the rats choice of pathways upon the length of the daily trial series. *J. comp. Psychol.*, 1934, 18, 135–147.
12. Brozek, J. Personal factors in competence and fatigue. (In preparation)

STUDIES IN SPATIAL LEARNING. III. TWO PATHS TO THE SAME LOCATION AND TWO PATHS TO TWO DIFFERENT LOCATIONS

BY BENBOW F. RITCHIE[1]

University of California

Introduction

In the first article in this series (1) evidence was presented that training rats over a specific path to food produces a disposition in these rats to short-cut to the food location, if the training path is blocked and a series of other paths is provided. This acquired disposition in the rats was called 'an expectation of food in the given location.'

In the second article of this series (2) evidence was presented that it is easier for rats to learn to go from two different starting paths to the same place in the room, than to learn always to make the same turn, say left, from two different starting paths.

The question raised in this third article is this: If rats are trained from two different starting paths, will learning be more rapid when these two paths lead (1) to the same goal place or (2) to two respectively different goal places?

Because of the results reported in the first two articles it seemed probable that learning to go to the *same* goal place from two different starting paths would be the simpler problem, and therefore would be learned more rapidly. That is to say, the rat may as a result of training on one of the two starting paths acquire a disposition to short-cut not only from that starting position, but also from the other starting position as well. If this were the case, then training from either one of these two starting positions might strengthen the tendency to take the correct route to the common goal place from the other or opposite starting position. But in the case in which the two starting paths lead to two different goal places, training on one of these two routes should have no influence on the tendency to take the other route. In short, in this latter case, the rat must acquire two separate spatial expectations, one of food in location L_1, and the

[1] This investigation is part of a co-ordinated research program of the Psychological Laboratory at the University of California. The writer is deeply indebted to Professor E. C. Tolman for generous advice and aid. The present paper is a portion of a dissertation presented to the Faculty of the Graduate School of the University of California in partial fulfillment of the requirement for the degree of Doctor of Philosophy.

other of food in location L_2. Each of these two dispositions would have to be acquired independently of the other, and thus more training would be required before the performance on these two routes would be perfected.[2] The following experiments were performed to test this hypothesis.

Experiment I

A. Animals

Sixteen pigmented male rats, approximately three months old, were used in this experiment. These rats were $M \times M$'s from the Tryon strains at the University of California laboratory. Four days before the beginning of this experiment, they were put on a 24-hour wet-food maintenance schedule. None of these animals had had any previous experimental training. They were fed every night at 10 P.M., and all the experimental trials were run between 8 and 10 P.M. When the experimental training began the 16 rats were separated into two groups of eight each. The group which was required to learn two paths to the same place was called 'the experimental group.' The group which was required to learn two paths to two different places was called 'the control group.'

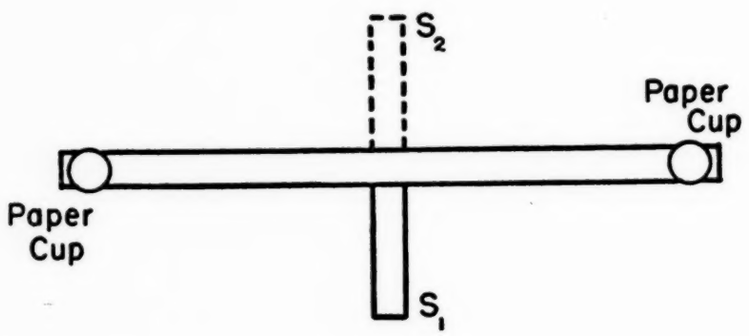

Fig. 1. Elevated maze. S_1 and S_2: alternative starting points

B. Apparatus

The apparatus used in this experiment (*cf.* Fig. 1) was an elevated maze similar to that used in the second experiment of this series (2), except that instead of having end-boxes and curtains at the ends of the center path, small paper cups, two in. in diameter and 1½ in. deep, were placed on both ends of the center path. The center path was two meters long, while the starting path was 40 cm.

The only illumination in the room came from an overhead frosted globe containing a 200 watt bulb (*cf.* Fig. 2). This lamp was suspended from the ceiling approximately seven feet above the maze. The animals lived in home cages kept in a rack which stood against the north wall. During all the training trials the animals were taken individually from their home cage in the rack and placed on the maze. After their trial they were then returned to a cage which rested on a stand in front of the home-cage rack (*cf.* Fig. 2), approximately two feet from the west wall. When all four of the rats in a given home-cage had been run, they were taken from the cage on this stand, and returned to the home cage in the rack to wait for the next trial.

The various positions of the maze in the room are indicated in Fig. 2. Position *A* was that employed during the preliminary training of both groups when position habits were being tested. Position *A* was also the position for the experimental group. During the experimental trials, positions *B* and *C* were the two alternative positions of the maze for the control group during the

[2] This argument is worked out in some detail in the discussion at the end of the present article.

experimental trials. The animals in this latter group were run two trials each night at position *B* and two at position *C*, while the experimental group ran all four of their trials in position *A*.[3]

During each of the trials the *E* sat on a stool directly behind the starting path, and put the rats on this starting path by hand.

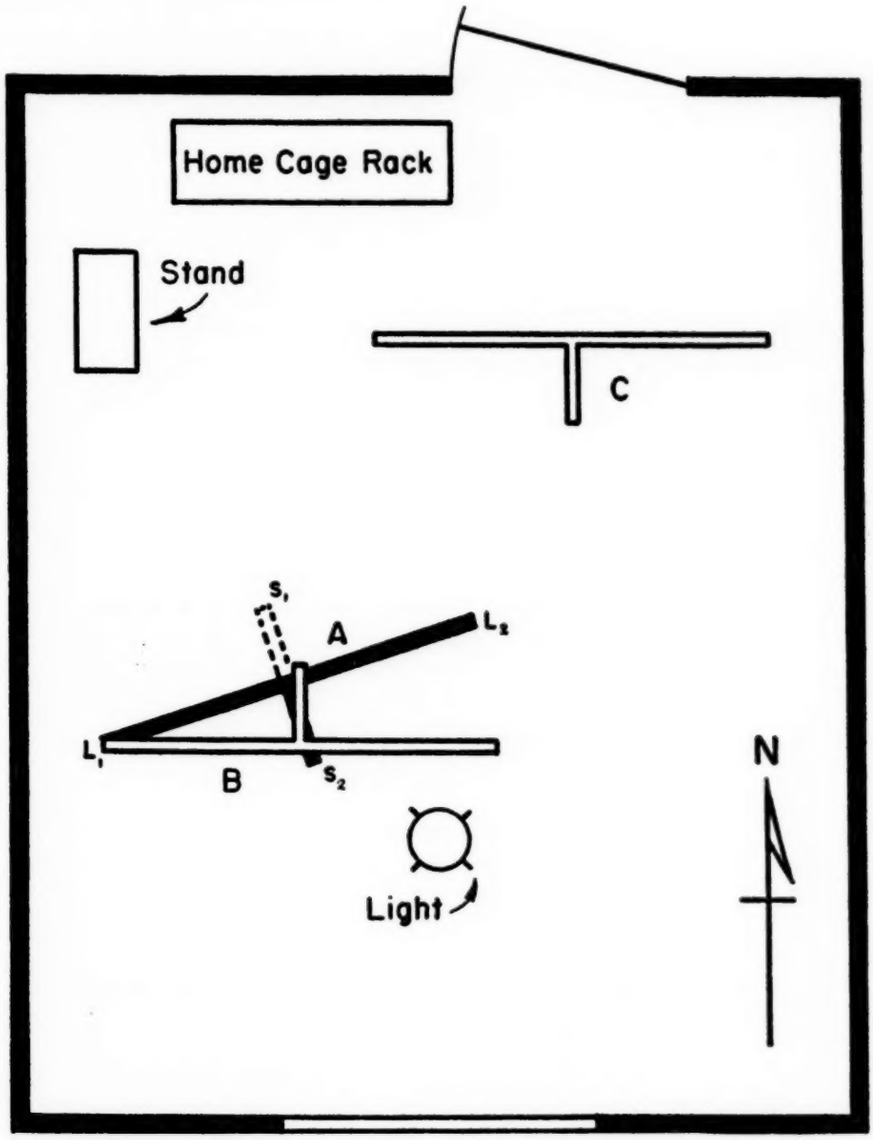

FIG. 2. Floor plan of room with mazes *A*, *B*, and *C*

C. Method

1. *Preliminary training.*—After three days on a 24-hour wet-food maintenance schedule, both groups of rats were given one day's training to accustom them to run on elevated paths and to eat from paper cups on the end of such paths. This training was given during the afternoon so that the lighting would be distinct from that used later in the experimental trials. Four elevated paths were set in the room so that they were parallel to each other and ran from north

[3] For the experimental group two of these trials were begun facing south (i.e., at S_1) and two were begun facing north (i.e., at S_2). Thus they were required to make two right and two left turns every night. For the control group all the trials in position *B* were begun facing south, and all those in position *C* were begun facing north. Thus, the control group was also required to make two right turns and two left turns every night.

to south. They were approximately three feet apart. Each of these paths was five feet long and had paper cups containing wet food at both ends. Each of the animals was given five 'trials' on these paths. Such a 'trial' consisted in being taken from the home-cage, placed in the middle of one of the four paths, allowed to run back and forth, and to eat from either or both of the cups for five min. Thus, at the end of this part of the preliminary training each rat had had 25 min. on some one of these four paths.

On the following night both groups were given four trials on the apparatus in position *A* (*cf.* Fig. 2) in order to determine their position habits. Two of these trials started from S_1 and the other two from S_2. The order of trials was $S_1S_2S_2S_1$. Food was placed in the cups at both ends of the center path, and the animal was allowed to remain on the path for 25 sec. after having first reached one of the two food cups. Some of the animals on some trials spent all their 25 sec. in eating from the same cup. At other times they would eat a little from one and then turn about and run to the other cup and eat there. These 'retracings' were recorded along with the hesitation times which preceded each choice.

One animal in each group exhibited a consistent right-turn habit, while two animals in the experimental group and three in the control group consistently chose a given side and this was later made incorrect. The other animals made no consistent choices.

2. *Experimental training.*—(*a*) Experimental group ($N=8$). The animals in this group were given four trials a night for four nights on the apparatus in position *A*. The order of starting positions was $S_1S_2S_2S_1$ on odd-numbered nights and $S_2S_1S_1S_2$ on even-numbered nights. The training procedure was the same as in the preliminary position habit test except that food was placed only in the cup at the west end of the path. Errors, VTE's, times, and retracings were recorded for each animal.

(*b*) Control group ($N=8$). The animals in this group were given the same number of trials per night for four nights. Each night they were given two trials with the apparatus in position *B* and two trials in position *C*. The order was *BCCB* on odd-numbered nights, and *CBBC* on even-numbered nights. The training procedure was the same as that for the experimental group with food only in the cup at the west end of each path. Errors, VTE's, times and retracings were recorded for each animal.

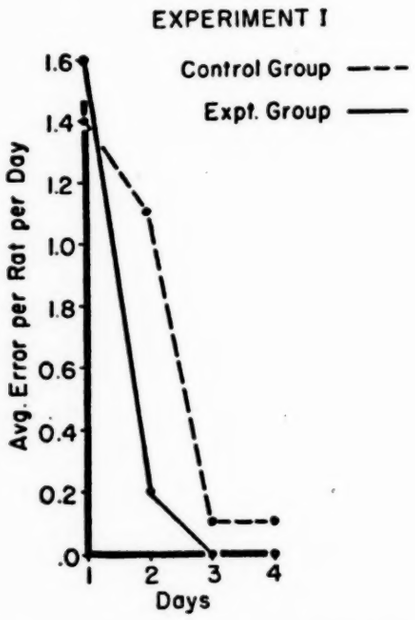

FIG. 3. Errors as a function of trials

D. Results

The progress of learning in the two groups is shown in the error curve in Fig. 3. The experimental group dropped rapidly on the second day from 1.6 errors per rat to 0.2 error on the second day,

while the control group only dropped to 1.1 errors per rat on the second day. However, on the third day the control group dropped to 0.1 error while the experimental group ran perfectly.

The mean number of errors per rat for the four days was 2.0 for the experimental group with a standard deviation of 0.7. The mean number of errors per rat for the four days was 2.8 for the control group with a standard deviation of 1.7. Consequently, the difference between the two groups was not statistically significant. The critical ratio of the difference between the means was only 1.15.

E. Discussion

The results of Experiment I indicate only a slight difference in favor of the experimental group, the group which learned two different paths to the *same* place. This difference was most marked on the second day of training. But after that, the difference between the two groups seems negligible and certainly is not statistically significant. Thus, we failed to find convincing evidence for our hypothesis that if rats are trained on two starting paths which lead to the same goal place learning is more rapid than it is when the two paths lead to two different goal places.

The question that now arises is, why did we fail to get more definite evidence? There are at least three alternative answers to this question.

1. Both groups may have learned on the basis of intramaze cues, since the center path was not rotated during the training. In this case we might expect the learning rates of the two groups to be the same.[4]

2. The experimental group may not have learned their problem as two routes to the same place, but merely may have learned when facing south to turn right, and when facing north to turn left. In other words, finding food at L_1 when having run from S_1 may not have increased the disposition to turn towards L_1 when started from S_2. The control group similarly may have learned to turn right when facing south and left when facing north. Thus, if this were the case, there would be no difference in the learning of the two groups, since both would be learning the same thing. This answer is simply a denial of the correctness of the hypothesis for which we were seeking evidence.

3. The final alternative answer is this: the control group may also have learned their problem as two routes leading to the *same* place. That is, there may have been some strong orientational cue in the

[4] Although, on this hypothesis, practice from one side of the maze would increase the disposition to take the same identical path (e.g., the one with the knot-hole) when started from the opposite side, this is not spatial learning in our sense of the term.

room which was the same for the mazes whether in positions *B* or *C*. If this were the case, then training on the maze in position *B* would increase the disposition to go in the same direction (i.e., west) when started on the maze in position *C*. Thus there would be no difference between the two groups, since both would be learning to go to the same place (or in the same direction). The *E* had hoped in designing this experiment to rule out this possibility, intending that the light should be the only orientational cue, and further he had hoped that the relation of the light to the goal when the maze was in position *B* would be distinct from its relation to the goal when the maze was in position *C*. There were, however, two possible slips in the carrying out of this intention: 1. The primary orientational cue may not have been the light, but may have been some other feature of the environment which the *E* had overlooked, and which was the same for the goal locations in both of the positions of the maze (i.e., *B* and *C*). Or, 2, the light may have been the primary orientational cue, but the rats may not have been able to distinguish the goal locations in terms of its relation to the light.

In order to discover if any of these three hypotheses was true, and thus to discover, if possible, why we failed to get the evidence for which we were seeking, further tests with the control group were carried out. During these tests food was put at both ends of the center path so that no further learning would occur.

On the fifth night the center path was turned end for end to determine whether the performance was based upon intra-maze cues. No change in performance resulted; only one error was recorded for the entire group in four trials.

On the sixth night the *E* tried to discover whether or not the light was an important orientational cue. To be specific, he tried to discover: 1, whether the rats had learned to turn right in the lighted end of the room and left in the dark end; or 2, whether they had learned to go towards certain specific places and away from others, these places being specified in terms of their relation to the light. If evidence were found that the rats had acquired *neither* of these dispositions, then it would be clear that the light was probably not an important orientational cue. To investigate the possibility that they had acquired niether of these two dispositions the maze was moved to the positions shown in Fig. 4*B*. (In Fig. 4*A* we again see the previous positions which had been employed until day 6.) Now if the rats had acquired either of these two dispositions, their behavior should break down (i.e., they should stop going west) when the maze was moved to these new positions. That is, if the rats were turning right in the light and left in the dark, and they continued this, they should all go east rather than west on these test trials with the maze in the

position shown in Fig. 4*B*. Similarly if they had learned to go to specific places and to avoid specific places in the room, they should go east when started at the south end of the room to avoid the place of the original blind, and they should also go east in the north end of the room in order to go to the place of the original goal. In conclusion, then, if the rats on these test trials broke down and began going east, we would have evidence that they may have acquired one of these two dispositions. *But, if they failed to break down, we would*

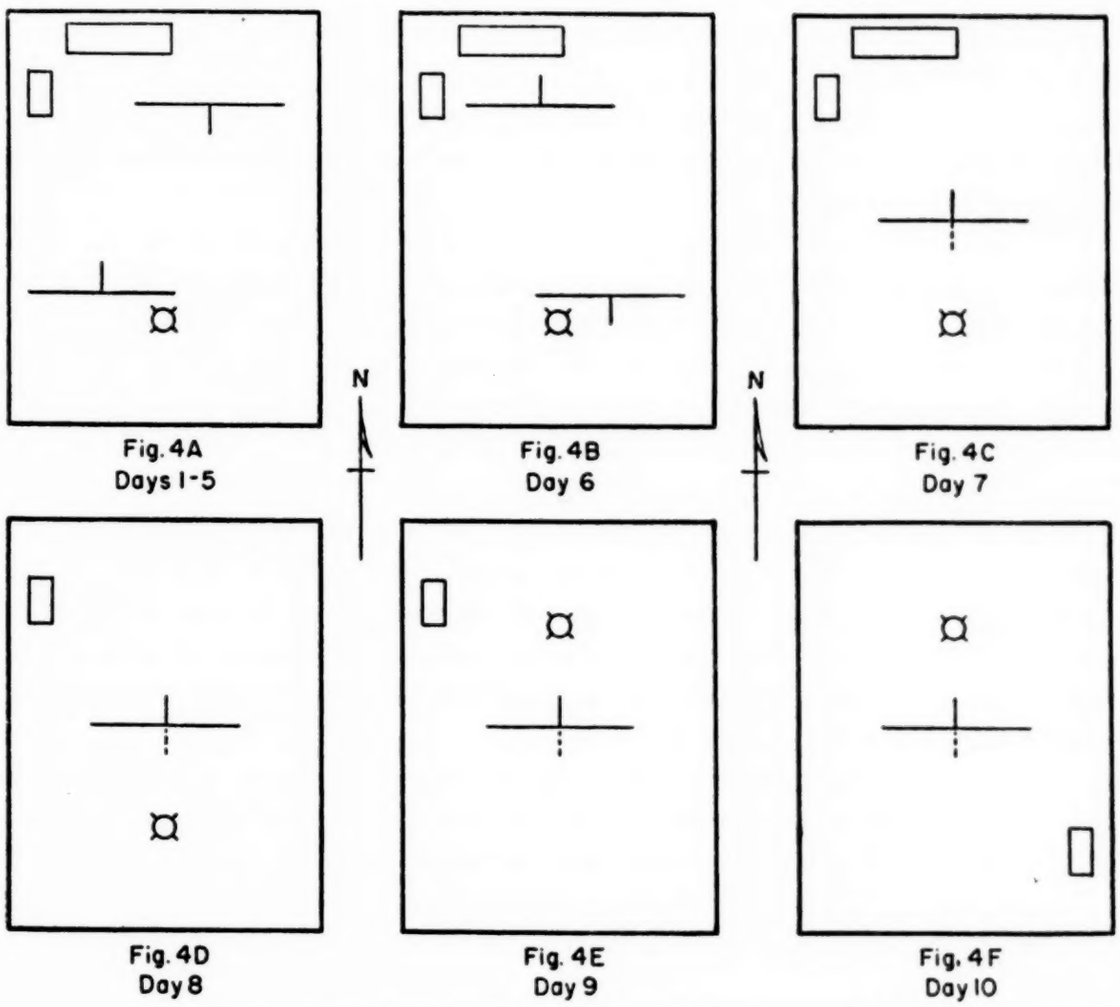

FIG. 4. Floor plans showing maze positions on successive days

have evidence that they had acquired neither of these dispositions, and that therefore the light was probably not an important orientational cue. When these test trials were run there was no significant change in performance. They showed no increase in hesitation, and only three errors (runs to the east) were recorded for the entire group in their four trails. This suggests that the influence of the light was slight.

On the seventh night the rats were run with the maze in the center of the room, as shown in Fig. 4*C*. This problem was similar to that learned by the experimental group. Only one error occurred

on this night. On the eighth night (Fig. 4*D*) the home-cages of all the animals were removed to another room, and a single cage of four animals was brought into the room for each trial. After the four rats in that cage had been given a trial they were returned to the other room, and another group of four rats was run.[5] This was done to determine the orientational influence of the home-cages. No change in performance resulted on the eighth night: one error and no increase in hesitation.

On the ninth night the maze remained in the same position, but the light was shifted in position as shown in Fig. 4*E*. This was a further and conclusive test of the influence of the light. If the light were the primary orientational cue then all the rats should go east on these trials. No change in performance resulted: four errors and no increase in hesitation.

At this point it became quite clear to *E* that there must be some strong orientational cue present in the environment which he had overlooked. He then hit upon the hypothesis that the cage which rested on the stand in front of the home-cage rack (*cf.* Fig. 2), from which the animals were carried before each trail and to which they were returned after each trial, was probably this overlooked cue. This cage stood only two feet from the west wall, and since the rats in it made some noise, it might very well serve as such a cue. In short, the rats in the control group may merely have learned to run towards this cage, and since the sound coming from it was diffuse and not easily localized, the goal location (defined in terms of this extra-maze cue) was identical for the rats in the control group whether the maze was in position *B* or *C*.

To test this hypothesis this stand and cage were placed on the 10th night in the opposite corner of the room (in the southeast corner) (*cf.* Fig. 4*F*). Each of the rats was given four trials with the cage in this position. Their change in performance was astounding. There was no increase in hesitation, but 24 errors out of a possible 32 were made.

In conclusion, then, the reason that we failed to get the evidence for which we were searching was that the control group learned their problem as two routes leading to the cage and stand. Thus, the experiment failed because we failed to set the correct kind of problem for the control group. Consequently, we decided that, if we could remove all the orientational cues but the light, and if we could make the two goal locations for the control group distinct, we should be able to get the evidence for which we were looking. With this end in mind Experiment II was designed.

[5] For all the following tests the home-cages were kept in this other room and the same procedure followed.

EXPERIMENT II

A. Animals

Fourteen pigmented male rats of the same age and stock as those used in Experiment I were used in this second experiment. They were put on the same maintenance schedule, and were fed every night at 10 P.M. None of them had had any previous experimental training. The one-place group ($N=6$) is called 'the experimental group.' The two-place group ($N=8$) is called 'the control group.'

B. Apparatus

The mazes used in the present experiment were exactly like the ones used in Experiment I, except that the center paths were 2.5 meters rather than 2 meters, and the starting paths were 50 instead of 40 cm.

Two rooms and two mazes were used for the control group, while only one room and one maze was used for the experimental group. The mazes were set in the rooms so that the choice points were in the exact center of the rooms. The only illumination in Room *A* (where the experimental group had all its trials) came from the same overhead lamp as was used in Experiment

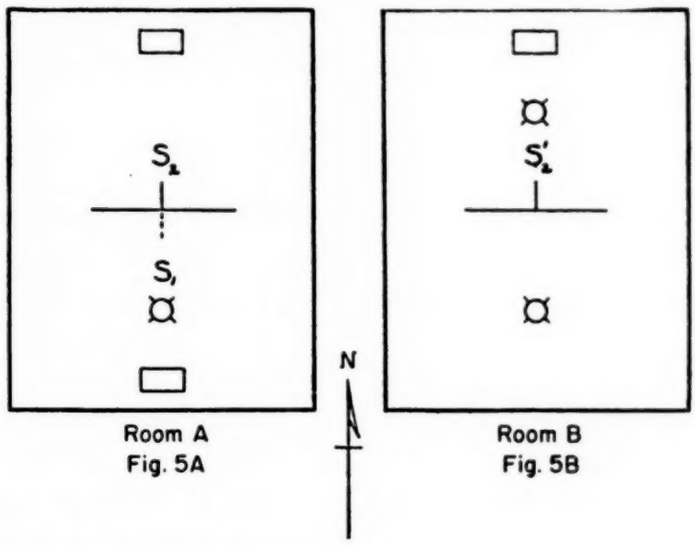

FIG. 5. Floor plans of rooms *A* and *B*

I (*cf.* Fig. 5*A*). The control group were given two trials a night in Room *A*, starting always from S_1, and two other trials in Room *B* from S'_2 (*cf.* Fig. 5*B*), where there were two overhead lamps along the axis of the starting path.

The animals lived four in a cage in still a third room and were carried during the experimental training trials from this third room to one of the experimental rooms, given one trial, and returned to the home-cage to wait for their next trial. The cage in which they were carried was always set exactly behind the starting path, in order to remove the cage cues which had compromised the results of Experiment I.

During each of these trials *E* sat on a stool directly behind the starting path, and put the rats on the path by hand. In addition the center path was turned end for end at the conclusion of the second trial in each night run of four trials. This was done to remove the influence of intramaze cues.

C. Method

1. *Preliminary training.*—After three days on a 24-hour wet-food maintenance schedule both groups were given three days' training to accustom them to run on elevated paths and to eat from paper cups at the ends of such paths. This training was given in the third or home-cage

room, where the lighting was distinct from that of either of the two experimental rooms. They were run four times a night on five-foot elevated paths which had food at only one end. During each run they were allowed to remain on the path for 60 sec. after they had run to an end and had begun to eat from one food cup. Thus, at the end of the three days of preliminary training each of the rats had had approximately 12 to 15 min. on elevated paths.

On the following night the animals were divided into two groups, the experimental, or the one-place group, and the control, or two-place group. The animals in both groups were given four trials in order to determine their position habits. The experimental group was tested on the maze in Room *A*, with food at both ends of the center path. Two of these trials started from S_1 and two from S_2. The order of trials was $S_2S_1S_1S_2$. The animals were allowed to stay on the path for 25 sec. after having reached one of the two food cups. The control group was tested two trials on the maze in Room *A* from the starting position S_1 and two trials on the maze in Room *B* from the starting position S'_2. The order of trials was $S'_2S_1S_1S'_2$, with food at both ends of both center paths.

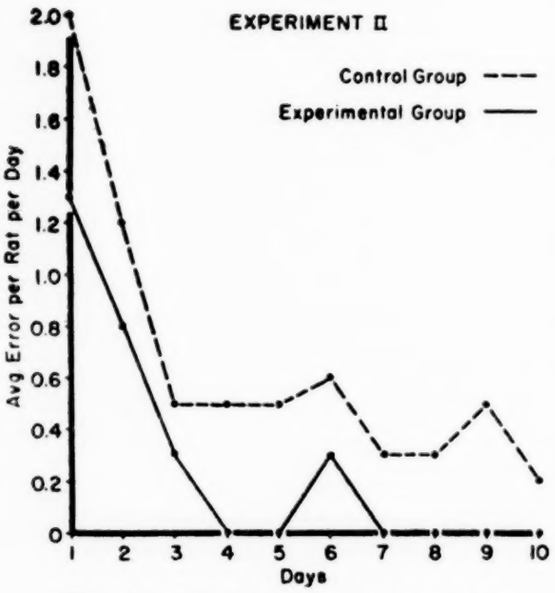

FIG. 6. Errors as a function of trials

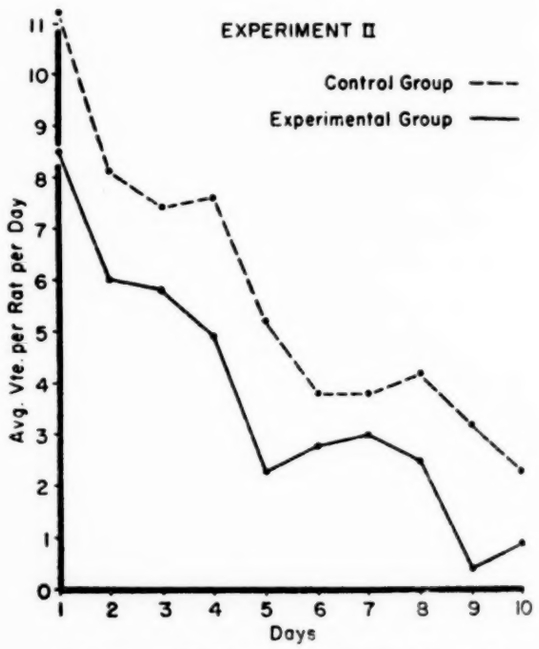

FIG. 7. VTE's as a function of trials

The results of this position habit test were as follows: Two of the 14 animals showed a consistent turn habit, both making four successive left turns, and both of these animals were in the experimental group. Two of the remaining animals showed a consistent direction habit (one going west, the other going east on all four trials), and both of these animals were in the control group. The remaining 10 rats made unsystematic choices.

2. *Experimental training.*—(*a*) Experimental group ($N=6$). The animals in this group were given four trials a night for 10 nights on the maze in Room *A*. The order of starting positions, and training procedure was the same as that for the experimental group in Experiment I.

(*b*) Control group ($N=8$). The animals in this group were also given four trials a night for 10 nights. Two of these trials were from S_1 in Room *A* and two from S'_2 in Room *B*. Thus, the training procedure was the same as it had been in the preliminary position habit tests, except that food was placed in only one of the two cups on the center path.[6]

D. Results

The progress of learning in the two groups is shown by the error curves in Fig. 6. Both groups dropped rapidly until on the third day,

[6] Half of the experimental group ran west on all trials and half ran east. The control group was treated similarly. Thus, the animals in both groups had on any given night to run either *RLLR*, or *LRRL* depending on whether they were in the east or west-going group.

when the control group reached a plateau at 0.5 of an error per rat. The experimental group, on the other hand, continued to improve until they ran four successive errorless days.

The mean number of errors per rat for the 10 days was 2.8 for the experimental group, with a standard deviation of 1.1. The mean number of errors per rat for the 10 days for the control group was 6.9, with a standard deviation of 4.3. We can see from these results that there was a statistically significant difference between the learning rates of these two groups. The critical ratio of the difference between the means was 2.37, which enables us to reject the null hypothesis at the two percent level of confidence.

The change in the amount of choice-point hesitation is shown by the VTE curves in Fig. 7. As these curves show, the experimental group consistently exhibited less VTE's at the choice point than the control group. By the end of 10 days' training, the experimental group was exhibiting less than one VTE per rat per day, while the control group was exhibiting more than two VTE's per rat.

E. Discussion

It is clear from the results of the second experiment that, in that situation at least, learning was more rapid when a rat was trained on two paths to the same place than when a rat was trained on two paths to two different places.

Before an hypothesis is presented to account for this fact two points should be emphasized.

1. The hypothesis which will be presented is one that can be directly tested itself.[7] This is an important point, since most of the so-called theories in contemporary studies of behavior are admittedly not directly testable, but can only be confirmed by testing theorems which are supposed to be logically or mathematically derivable from these theories. Our hypothesis, on the other hand, is a statement about the relation between certain training procedures and the dispositions believed to result from these procedures. Such an hypothesis obviously admits of direct confirmation. This is a tremendous advantage, since the hypothesis can be confirmed or rejected by the performance of a single experiment, provided the latter is appropriately designed. But hypotheses which are not directly testable can only be tentatively accepted after a great deal of indirect evidence has been accumulated.

2. None of the hypotheses or 'theories' current in psychology today seem to account for the results which we have reported, unless some should argue that the problem which was set for the rats was a

[7] Later papers in this series will seek to test this hypothesis directly.

complex discrimination. That is, according to such a notion, the rats in the experimental group were required to learn to turn right when facing the light and left when the dark was behind them. The rats in the control group, on the other hand, were required to turn right in the room with one light and left in the room with two lights. But, although such a suggestion seems a reasonable one, it offers no explanation for the slower learning in the controls. In fact, according to this suggestion we should expect that the experimentals would be slower learners, since the differential cues presented to them were more alike than those presented to the controls. None of the other current hypotheses seem applicable.

Now let us turn to our explanation of the results. First, let us give a rather loose common sense statement of our hypothesis which we can restate later in a more precise fashion.

The common sense statement would be: no matter where they are started, the rats in the experimental group always find food in the same place. As a consequence they come to expect food in that place. The rats in the control group, on the other hand, find food sometimes in the one place and sometimes in the other (although it is always in the same place in a given room). Thus, the control group come to expect food in two different places. Now let us consider the difference in the two groups when at the choice point and faced with a choice between two paths. The rats in the experimental group have merely to choose which of the two paths lead to the expected food place. But the rats in the control group have to decide (1) where in the given room to expect the food, and then (2) which of the two paths lead to this expected food place. Thus, there are two sources of error for the control group, while there is only one source of error for the experimental group. Because of this, it seemed probable that the problem for the control group would be the more difficult. But once the controls learn where to expect the food in each of the two rooms, and learn to choose the paths to these expected food places, they then can perform the problem as well as the experimentals.

This is our hypothesis. Let us now attempt to state it more precisely by making clear what we mean by 'expectation' as a disposition to short-cut to the food place.

In the first article of this series (1) we presented evidence that training on a specific path to food (*cf.* Fig. 8*A*) produces a disposition to short-cut to that food place (*cf.* Fig. 8*B*) when the trained path is blocked. This much of our explanatory hypothesis we know to be true. Now let us consider what we further believe to be true, although we have at present no evidence for this belief. These further beliefs we will call 'assumption 1 and assumption 2.'

Assumption 1. We further believe that training over a specific path produces a disposition to short-cut not only from the original starting position but also from other positions in the room. Thus, if we trained rats over the path illustrated in Fig. 8*A* and then tested them on the short-cut problem from the other side of the choice point (*cf.* Fig. 8*C*), we believe that they too would short-cut to the food place.

Assumption 2. We also believe that if a rat has a disposition to short-cut from S_2 to F (as in Fig. 8*C*), it will when started at S_2 (as in Fig. 8*D*) tend to choose the path leading to the food place rather than the path leading away from it, although it may never have been trained on the former path.

Now let us consider what differences there would be between the experimental and control groups if these two assumptions were true.

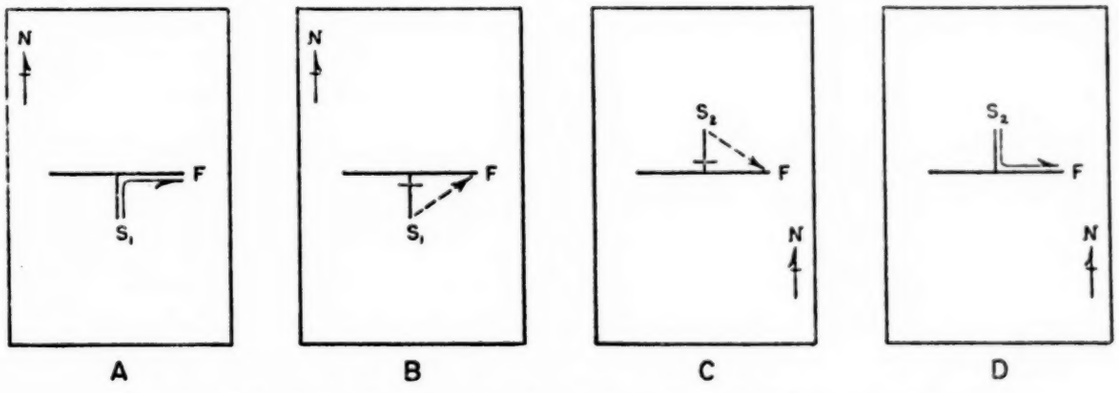

FIG. 8. Schematic diagrams showing assumed relations between short-cut dispositions and choice point turns.

According to assumption 1, training from S_1 to F produces a disposition to short-cut to F from S_2 as well as from S_1. And, according to assumption 2, rats with a disposition to short-cut from S_2 to F, will when started at S_2 tend to turn towards F when they reach the choice point. The more training they have in running from S_1 to F, the stronger will be the disposition to short-cut from S_2 to F, and the stronger will be the tendency to turn towards F when at the choice point.[8] In other words, according to our hypothesis there is a transfer of training from one side, as say S_1, to the opposite side, S_2, so that each run from S_1 strengthens the tendency to run to the food place from S_2 as well as from S_1.

In the case of the controls the situation is quite different. Here training from S_1 in Room *A* produces no disposition to short-cut from S_2' in Room *B*. As a consequence there is no transfer of training from S_1 to S_2' and each of the 'sides' has to be learned separately. From this it follows that the problem for the control group is more

[8] See article IV in this series for a qualification of this statement.

difficult than the problem for the experimentals. The results which we have reported confirmed this prediction.

It is in this precise sense that the concept of 'expectation' explains the results of our experiment. It should be clear that this explanatory hypothesis is not *ad hoc*, since it may very well be false, in the sense that assumptions 1 and 2 may turn out upon experimentation to be false. If this proves to be the case, then we must abandon the expectation hypothesis (at least in its present formulation) and find some other principle which will explain these results and which can itself be directly tested. Only in this way can one avoid the *ad hoc* fallacy which infects so much psychological 'theorizing.'

Finally, it is clear that the hypothesis presented seems to explain the facts which have been discovered. The only question now is whether this hypothesis is true. To determine this, one must test the two assumptions which were stated in formulating the hypothesis. This will be attempted in later papers in this series.

F. Summary

1. In Experiment I two groups of rats were trained on comparable single unit maze problems. The experimental group was required to learn to go from two starting paths to the same food place, while the control group was to learn to go from two paths to two respectively different food places.

2. No significant difference in the learning rates of the two groups was discovered. The reasons for these results were investigated and it was discovered that the orientational cues had not been adequately controlled.

3. In Experiment II the orientational cues were better controlled and two new groups of rats were run in the same type of problem.

4. At the end of 10 days from Experiment II the mean number of errors per rat in the control group was 6.9 with a standard deviation of 4.3. The mean number of errors per rat in the experiment group was, however, only 2.8 with a standard deviation of 1.1. The critical ratio of the difference between the means was 2.37.

5. An hypothesis based upon the notion of 'expectation' as a disposition to short-cut to the food place was suggested as an explanation of these experimental results. This hypothesis was one which can be directly tested and thus suggests further experimentation.

(Manuscript received February 25, 1946)

References

1. Tolman, E. C., Ritchie, B. F., & Kalish, D. Studies in spatial learning I: Orientation and the short-cut. *J. exp. Psychol.*, 1946, **36**, 13–24.
2. Tolman, E. C., Ritchie, B. F., & Kalish, D. Studies in spatial learning II: Response learning versus place learning. *J. exp. Psychol.*, 1946, **36**, 221–229.

STUDIES IN SPATIAL LEARNING. IV. THE TRANSFER OF PLACE LEARNING TO OTHER STARTING PATHS

BY E. C. TOLMAN, B. F. RITCHIE, AND D. KALISH[1]

University of California

Introduction

In the third article of this series (2) it was discovered that it was easier for rats to learn two paths to the same place than to learn two paths to two different places. In the argument which suggested that such results would be obtained was an implicit assumption that training on one starting path to a given food location would facilitate learning to run to this same food location from another starting path. That is, it was assumed that, if rats were trained from S_1 to L and then were put on S_2, a place they had never been before, they would perform from S_2 better than rats which had not had such preliminary training. It is this assumption which the present experiments were designed to investigate.

Experiment I

Animals

The animals used in this experiment were $M \times M$, pigmented, male rats from the Tryon strains maintained in the Psychology Laboratory at the University of California at Berkeley. At the beginning of the experiment their ages ranged from 120 to 150 days. All of the animals were maze-wise. Some had been previously run in an experiment reported by Kalish (1); the others had been previously run in the experiment reported in the third article of this series (2). In neither of these experiments was the problem such as to provide either positive or negative transfer to the present one. Each of the groups run in this experiment had 50 percent of its animals drawn from each of the two sources.

Apparatus

A simple elevated T-maze with a starting path section 50 cm. long and two arms each 125 cm. long was placed in the north-west end of a room 4 meters wide, 5 meters long, and 3½ meters high (*cf.* Fig. 1). At the end of each arm of the maze was a small paper cup which served as a food cup. Besides the maze the room contained nothing but three boxes, two for supporting cages and one for supporting a desk lamp, a chair for the *E*, and the desk lamp itself which was the sole source of light. This lamp was 2½ meters off the floor in the north-east end of the room; it faced south-west and cast a diffuse light in that direction. This directional lighting placed the end of the east arm of the T directly in the light, and the end of the west arm relatively in shadow.

Procedure for Group I (Experimental group, $N=7$)

Four days before an animal was run in the experiment he was put on a 24-hour wet-food maintenance-schedule, the feeding hour being ten o'clock in the evening. The experimental

[1] The cost of this investigation was met in part by grants to the Department of Psychology from the Research Board of the University of California.

trials were run between eight and ten in the evening. The previous training in the other experiments in which the animals had been used was similar enough to that given them in the present experiment to eliminate the necessity of much preliminary training to familiarize them with handling, running of elevated paths, etc. However, one day of preliminary training was given to determine each rat's position habit. This was done by running the animals six trials with food

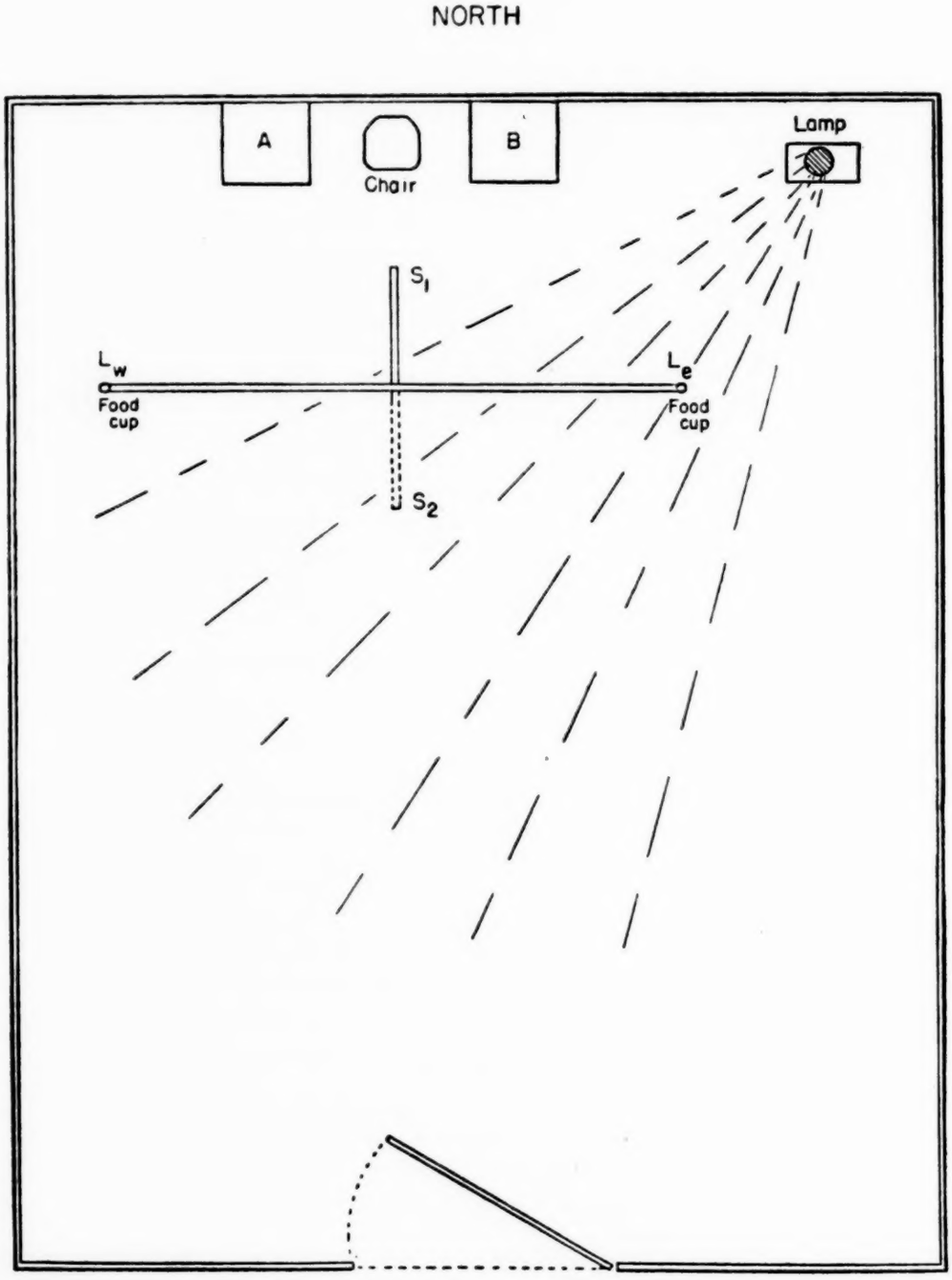

FIG. 1. Plan of room and apparatus

at both ends of the maze. The starting path was on the north side of the arms of the T during this training (*cf.* Fig. 1). In terms of their behavior on this day the rats were classified into two groups: (1) right-turn (west-going) rats, and (2) left-turn (east-going) rat . Those rats which went three times to each end were arbitrarily divided between the two groups.

The day following that on which position habits were determined the experimental trials proper were started. The starting path was, as on the preceding day, on the north side of the

arms of the T. The problem was set so that those animals that were in the right-turn (west-going) group found food on their left (east), and vice versa for the animals in the other group. Five trials, correction method,[2] with 25 sec. for feeding[3] were given to the animals. In order to eliminate uncontrolled spatial orientation cues given by living cages (*cf.* 2), a cage containing four (or three) rats was brought from the room in which the animals lived into the experimental room. (As indicated above, no animals lived in the room in which the experiment was performed.) This cage was placed either on box A or box B (*cf.* Fig. 1), alternating boxes each night; the other box supported an empty cage. The *E* sat in a chair between boxes A and B and directly behind the starting path.

After an animal had completed his trial he was removed from the apparatus and placed in the cage which he had not been taken from. An animal was not given another trial until the other three (or two) animals in that cage had had their trials. When the animals of one cage had finished their training for the day, they were returned to the home room before the next cage of animals was brought into the experimental room. It should also be noted that by interchanging the arms of the T at various times throughout the experiment the intra-maze cues were varied in a random manner.

After the fifth trial the starting path was immediately carried around the maze and attached to the south side of the arms of the T. Then a sixth trial was given from this, the opposite, side. On this trial the animal was carried by the *E* around the apparatus, but not always around the same end, and placed on the starting path. After the *E* had placed the animal on the starting path, he stood directly behind it throughout the trial. The non-correction method was used on this trial. That is, when an animal ran to the wrong end of the maze he was not allowed to correct his error; instead he was picked up by the *E* and immediately placed back at the starting position. This was continued until the animal made a correct choice. After the animal had made a correct choice he was returned to the cage from which he had not been taken. This whole procedure was counted one trial.

Procedure for Group II (Control group, N=7)

The control experiment consisted of running a group of seven rats with the same procedure as that of the experimental group, except that trials one to five of each day were run in another room on straight paths with food cups at one end. After their five trials to food, which gave them the same amount of exercise and food as that had by the experimental group, the rats were carried immediately into the room in which the experiment was performed and given their sixth trial on the simple T-maze. This test trial was identical in procedure to that given to the experimental group.

Results

For Group I, the experimental group, the mean number of errors per rat for 10 days (50 trials) of training on the practice trials (that is, trials 1 to 5 of each day) was 1.85 errors, with an SD of 1.6 errors. The mean number of errors per rat for 10 days (10 trails) of performance on the crucial test trial of each day for this group was 2.0, with an SD of 2.4. For Group II, the control group, the mean per rat for 10 days (10 trails) of performance on the test trail was 2.0 with an SD of 1.8. These results are shown in Table I.

[2] The correction method is contrasted here with the non-correction method. (*Cf.* Kalish, 1.) When trained by the former method rats are allowed to correct an incorrect choice; that is, if an animal turned right at the choice point when left was correct he was allowed, after having reached the end of the blind, to return to the goal, at the opposite end of the maze and eat for 25 sec. When trained by the non-correction method the rats were not allowed to correct an incorrect choice; that is, if an animal made an incorrect turn he was removed from the apparatus immediately after he reached the end of the blind and started again until he finally chose the correct side. Not until he had chosen the correct side was the trial considered complete.

[3] Some animals did not eat continuously for 25 sec., however, 25 sec. after an animal had first taken some food he was removed from the apparatus.

TABLE I

RESULTS OF EXPERIMENT I

	Group I ($N=7$)		Group II ($N=7$)
	50 Practice Trials	10 Test Trials	10 Test Trials
Mean Errors	1.85	2.0	2.0
SD	1.6	2.4	1.8

Discussion

Inspection of Table I shows that there is no difference between the performance of the animals in Group I on the crucial trial and the performance of the animals in Group II. Thus, the results of Experiment I are negative evidence for the assumption being investigated, at least in the present set-up, where the two starting-places in question are opposite starting legs of a simple T-maze. In other words, we started with the statement form:

> (A) If rats are trained from S_1 to L and then put on S_2, a place they have never been before, they will perform from S_2 better than a group which has not had the preceding training[4].

Now, our experimental set-up determines the values of these variables. S_1 becomes the starting leg on the north side of the arms of the T. S_2 becomes the starting leg on the south side of the arms of the T. L becomes the goal location at the end of the west (or east) arm. And the training is that described for Experiment I. Thus, our experiment tests directly the hypothesis:

> (B) If rats are trained (in the given room and with the given lighting conditions, etc.) from a starting leg on the north side of the arms of a T-maze to food at the end of the west (or east) arm of the T and then are put on a starting leg on the south side of the T, they will perform from this second starting position better than a group which has not had the preceding training[4].

And the results of our experiment show that this hypothesis is false.

Most false hypotheses are as interesting as true ones. This is the case if for no other reason than that an understanding of why an hypothesis is false often suggests other hypotheses which are true. First let us ask what it is that we may assume that the rats learned while performing in Experiment I. The training they received on trials 1 to 5 of each night had two salient characteristics: (1) a left (or

[4] The statement form A becomes an hypothesis only when the variables 'S_1,' 'S_2' and 'L' are bound or replaced by the names of specific starting places and a goal location. To bind the variables is to assert an hypothesis for any choice of a pair of starting places and a goal location. But such an hypothesis could be known to be true only if we had investigated each choice of a pair of starting places and a goal location. Hence, it is with such choices of a pair of starting places and a goal location that we are here concerned.

right) turn took them directly to food, whereas the opposite turn led them to food by a more devious route, and (2) the food was always in the same direction with respect to the light. Thus, both features were probably involved in their learning. That is, being in a hurry to get to food they developed a tendency both to turn left and to turn toward the direction of the shadow (or right and toward the direction of the light). However, when on the sixth trial of each night's training they were started from the opposite side of the maze, they were presented with a conflict. The two features upon which their learning had been built were now diametrically opposed; that is, if the animal were to continue to make the same turn with respect to the light (or shadow), he had to reverse the right handedness or left handedness of his turn, and vice versa. Hence, they had conflicting tendencies which manifested themselves in their performance.

Let us consider in detail the behavior of rat number 105. It appeared that for this rat the turns he made on trials 1 to 5 (right-hand turns) had a predominant effect. For on the sixth trial when a right turn led into the light instead of into the shadow he continued to exhibit a 100 percent tendency to turn to the right. This rat never made an error on trials 1 to 5 (suggesting that he probably had a native predilection for right turns), and after 10 days of training he was still making an error on the crucial trial (i.e., continuing to turn right). On the eleventh day a new prodecure was introduced for this one rat. He was given five preliminary trials similar to those given the animals in Group II. That is, he was run in another room five trials on a straight path to food. Then, he was given his sixth trial on the apparatus in the ordinary manner. On this trial he performed correctly. This procedure was continued two more days, and on both these further days he performed correctly on the crucial trial. On the next day, he was trained as in the original experiment, and now performed correctly from both sides.

The above discussion suggests an explanation of the negative results of Experiment I. In learning the simple T-maze problem the rats developed either (1) a tendency to go in a certain direction or (2) a tendency to make a certain turn or (3) both. Those animals that developed a tendency to go in a certain direction would be able (if they utilized this tendency) to perform correctly on the crucial test trial. Those animals that developed a tendency to make a certain turn would not be able (if they utilized this latter tendency) to perform correctly on the crucial trial. Those animals that developed both tendencies would have to learn to inhibit the turn tendency, if they were to perform correctly on the crucial test trial. We would expect only those animals that developed solely a tendency to go in a certain direction to perform better on the crucial test trial than

animals without any relevant preceding training. The animals with a turn tendency would be inferior, and the animals with both tendencies would have a problem which would be comparable in difficulty to that of the original learning of a simple T-maze.

This analysis fits in perfectly with our results. It seems probable that most animals developed both tendencies, and hence we should expect to find their behavior on the crucial test trial comparable to that of a group which learned the problem without the advantages or disadvantages of relevant preceding training. And this is what we did find. The behavior of the one animal, rat number 105, whose performance was definitely inferior to that of the animals of Group II can be explained by assuming that he was predominantly a turn animal. This assumption is justified both (1) by the fact that he made no errors in the training from the north side, and (2) by the fact that he performed correctly as soon as the frequent practice of the right turn was eliminated. This second fact, that he performed correctly as soon as the frequent practice of the turn was omitted, indicates that he was not totally unaware of the goal location, but that his performance on the crucial test trial, as long as the turn practice was present, was entirely independent of any special orientational factors. However, as soon as the turn practice was removed, he did perform in terms of spatial direction. And once he was performing in this manner, the re-insertion of the turn practice did not upset his behavior.[5]

The foregoing analysis of the negative results of Experiment I suggested an experimental set-up which ought to give positive results. The inhibiting factor of the crucial test trial was, we concluded, the opposing turn. Hence, any set-up in which the correct response from the new starting place is not a turn directly opposed to that which was correct from the original starting place would obviate this inhibiting factor. Experiment II was designed in this manner.

Experiment II

Animals

The animals used in this experiment were maze-wise rats from the same stock and with the same previous training as those used in Experiment I.

Apparatus

The apparatus was the same as that used in Experiment I, and it was located in the same room under the same environmental conditions.

[5] To investigate further the effect of turn practice on performance, the animals that were performing correctly on the crucial trial were subsequently given 10 instead of 5 trials of preliminary practice from the north side before being given their crucial trial from the south side. This increase in turn practice did not break down their correct performance on the crucial trial. Hence, we can conclude that *learning* from the south side was inhibited by practice of the opposite turn from the north side, but once the animals were performing correctly from both sides, then their subsequent *performance* was not affected by increased practice of the opposing turn.

Procedure for Group III (N=8)

The preliminary training for Group III was identical with that given to the animals in Experiment I. However, they were not grouped according to position habits in the same manner as the animals run in that experiment. The eight rats of Group III were divided into two groups of four each, each group containing some animals that had their position habit favored by the required turn and some that did not. That is, of the group that was to find food at the end of the west arm of the T, one of the animals had turned west, i.e., right, on the majority of preliminary trials, two had turned east, i.e., left, on the majority of trials, and the fourth animal had shown no preference. The group that was to find food at the end of the east arm of the T had a comparable assortment.

The day following the one on which position habits were determined the experimental trials proper were started. Trials 1 to 5 of the first day of training were run without the starting leg being attached to either side of the arms of the T. That is, the animals of the group which found food at the end of the west arm of the T were placed on the end of the east arm and allowed to run down the straight path to food at the west end, and vice versa for the other group. Everything else about the handling and training, including the random variation of intra-maze cues, was done exactly as it was for Group I in Experiment I.

After the five trials of running from the end of one arm to food at the end of the other, the animals were given a sixth trial identical in procedure to that given to the animals of Group I in the first experiment. That is, the starting leg was attached to the south side of the arms of the T, and a trial (non-correction method) was given on the resulting T-maze. This procedure was continued four days. On the fifth day of training the sixth trial was run from the north instead of the south side of the arms of the T. And on the sixth day, the five trials on a straight path to food were omitted, and four trials, the first with the starting leg on the north, the second and third with it on the south, and the fourth with it again on the north, were given.

Results

On the first day of training one rat made an error on the crucial test trial. There were no errors on days two, three, and four. The mean errors and standard deviation for the rats of this group on the crucial trials for the first four days of training are presented in Table II along with those for Group II of Experiment I. On the fifth day

TABLE II

RESULTS OF EXPERIMENT II

	Group II (*N*=7) (10 Test Trials)	Group III (*N*=8) (4 Test Trials)
Mean Errors	2.0	0.125
SD	1.8	0.33

of training, the day on which the sixth trial was run from the north instead of the south side of the arms of the T, they again all performed correctly. On the sixth day, that on which they were given four crucial trials without preceding preliminary practice, two from each side of the arms of the T, seven out of the eight rats performed correctly. One of the animals ran at chance.

Discussion

The behavior of the animals in Group III on the first four days of training as compared with that of the animals of Group II of the first experiment provides unambiguous evidence that the assumption being

considered is true for at least one set of values for the variables. That is, the experimental set-up of Experiment II determines S_1 (trained on starting place) as the end of the east (west) arm of the T-maze, S_2 (new starting place) as the end of the starting leg attached to the south side of the arms of the T, and L (goal location) as the end of the west (east) arm of the T. Thus, the experiment investigated the hypothesis:

> (C) If rats are trained (in the given room and under the given lighting conditions, etc.) from either the east or west end of the central path of a T-maze to food at the opposite end and then are put on a starting leg on the south side of the arms of the T, they will perform from this second starting place better than a group which has not had this preceding training on the central path.

And the results of the experiment show that this hypothesis is true.

The performance of the animals of Group III on the fifth night of training indicates that the end of a starting leg attached to the south side of the arms of the T is not the only place from which they were able to perform better than a group that did not have the relevant preceding training. Their correct performance when the starting leg was attached to the north side of the arms of the T also reinforces the argument that the negative results in Experiment I can be attributed to the frequent practice of a 'turn' which becomes incorrect on the crucial test trial. The animals that performed correctly from the starting leg on the north side of the arms of the T the first time they were placed on that starting leg had made the opposing turn to food from the starting leg when it was attached to the south side of the arms of the T. But they had performed the opposing turn only once each night, and they had had no experience of this opposing turn on this night in which they first performed correctly from the north side. Hence, it seems very plausible that the turn practice was the inhibiting factor in Experiment I. And the fact that seven of the eight animals in Group III maintained their correct performance on the sixth night when they were given crucial test trials from both sides adds further weight to this conclusion. It appears from the performance on the fourth and fifth days that the animals had 'learned' to perform correctly no matter to which side the starting leg was attached, and hence it was to be expected that they would maintain their performance on the sixth day which necessitated the making of opposing turns.

Summary

1. An hypothesis of the form:

> (A) If rats are trained from S_1 to L and then put on S_2, a place they have never been before, they will perform from S_2 better than a group which has not had the preceding training.

was investigated for two different sets of values for the variables.

2. The first determination of the variables gave the hypothesis:

> (B) If rats are trained (in the given room and with the given lighting conditions, etc.) from a starting leg on the north side of the arms of a T-maze to food at the end of the west (or east) arm of the T and then are put on a starting leg on the south side of the T, they will perform from this second starting position better than a group which has not had the preceding training.

This was discovered to be false.

3. It was argued that the fact, that the *turn* required for the correct performance on the crucial test trial was directly opposite to that which was correct on the preceding practice trials, was the explanation of these negative results.

4. The second determination of the variable, made because of this suggested explanation of the negative results of the first experiment, gave the hypothesis:

> (C) If rats are trained (in the given room and under the given lighting conditions, etc.) from either the east or west end of the central path of a T-maze to food at the opposite end and then are put on a starting leg on the south side of the arms of the T, they will perform this second starting place better than a group which has not had this preceding training on the central path.

This was discovered to be true.

5. Analysis of the results of the second experiment lent support to the suggested explanation of the negative results of the first experiment.

(Manuscript received March 6, 1946)

References

1. Kalish, D. The non-correction method and the Blodgett and McCutchan delayed response problem. *J. compar. Psychol.*, 1946, **39**, 91–107.
2. Ritchie, B. F. Studies in spatial learning. III. Two paths to the same location and two paths to two different locations. *J. exp. Psychol.*, 1947, **37**, 25–38.

ON CERTAIN CONDITIONS CONTROLLING THE REALISM AND IRREALISM OF ASPIRATIONS

BY MALCOLM G. PRESTON, ANNE SPIERS AND JOYCE TRASOFF

University of Pennsylvania

In a recent theoretical paper, F. W. Irwin (1) is able to give experimental support to the distinction between realistic and irrealistic levels of aspiration. By realistic aspirations are meant those aspirations which, intelligently or unintelligently, are based upon an appraisal of the extent to which the individual is capable of meeting the demands of the situation with which he is confronted. An irrealistic aspiration, on the other hand, is one which is based upon the hopes, wishes, fears, and other factors, which in a certain sense may be attributed to the subjective states of the person yielding the statement of aspiration.

Certain deductions follow immediately from these definitions. For example, realistic aspirations will be functionally dependent upon performance, but not upon other realistic aspirations; i.e., statements of expectation should be correlated with measures of performance rather than with succeeding statements of expectation. Again, irrealistic aspirations, e.g., statements of hopes, will be functionally independent of performance and functionally dependent upon other such statements; i.e., the correlation of a hope with a hope will be higher than the correlation of a hope with a performance score. An important section of the Irwin paper is the demonstration that these deductions can be confirmed by supplementary analysis of experimental data already in existence.

A common technique for evoking aspirations of a given type of realism is the use of carefully formulated instructions. Thus, Irwin and Mintzer (2) succeeded in getting aspirations which were different in numerical value, depending upon whether 'hopes' or 'expectations' were required by the instructions. Now, while control of the instructions certainly represents the first step to be taken in the direction of controlling the realism of a stated aspiration, it does not follow that even well-formulated instructions will necessarily succeed in evoking aspirations of a stated degree of realism. In order to execute such instructions, the *S* must be capable of distinguishing between aspirations which genuinely are based upon objective reality, and aspirations which, in an equally genuine sense, are merely reflections of hopes and fears. Such a distinction is not easy to make, as is

frequently observed in games of chance. For example, in the game of poker the player occasionally must balance the hopes and fears engendered by the presence of a large prize against his evaluation of the probability of such an outcome as filling an inside straight. In any such case it is a question of fact, however thorny, whether a decision which the player states and believes to be based upon a cold, intellectual appraisal of the probabilities, was in any genuine sense, so founded. The ways of self-deception are legion.

In the level of aspiration experiment failure to realize the intent of the instructions may take various forms. Thus, an *S* asked to report an expectation may base it principally upon his hopes and fears, hence yielding an aspiration intrinsically irrealistic under circumstances where the instructions demand a realistic aspiration. Again, an *S* asked to report a hope, who has no identification with the task and hence no hopes nor fears, may found his statement principally upon his knowledge of the facts, thus yielding an aspiration which is intrinsically realistic under circumstances where the instructions demand an irrealistic aspiration. It was an objective of the present experiment to investigate the conditions under which stated aspirations may be identified as being genuinely realistic or irrealistic.

Three sources of experimental variation of realistic as well as irrealistic levels of aspiration were used in the experiment. One of these sources was the difficulty of the task; this was graduated in order to provide a basis for testing the extent to which stated aspirations were dependent upon performance. The other two sources were suggested by hypotheses concerning conditions under which such dependence would be found. The first hypothesis was that increase in the motivation of the *S* would intensify hopes and fears, and would thereby render irrealistic aspirations relatively independent of performance. Furthermore, since intensification of hopes and fears could not directly increase the extent of information about the requirements of the performance (a necessary condition for the sound development of a realistic aspiration), it was supposed that, with increase in motivation, statements of expectation would become relatively irrealistic, i.e., relatively independent of performance.

The second hypothesis was that increase of information about the requirements of the performance should have two consequences, an increasing realism of stated expectations and a sharper distinction between realistic and irrealistic bases of report. From this reasoning it was concluded that *S*s who were experienced in a task would give a sharper distinction between the extent to which their expectations were dependent on performance and the extent to which their hopes were independent of performance.

Readers who are familiar with the analysis of variance will recognize that the experimental questions here depend for their answer upon the nature of certain triple interactions. By the double interaction, aspiration × difficulty, is meant the extent to which two types of aspiration are similarly dependent upon difficulty. If the two types of aspiration exhibit differences in the way in which they depend upon difficulty, a double interaction is said to exist. Under the circumstance that no difference exists in the dependence of two types of aspiration upon difficulty when the *S*s are relatively ignorant, while a substantial difference exists in such dependence when the *S*s are informed, a triple interaction is said to exist. Formally, a triple interaction exists when a given double interaction holds under one condition but is absent under another condition. From the foregoing it is clear that the immediate object of the experiment was to provide a means for testing for the existence and properties of two triple interactions, i.e., Difficulty × Aspiraton × Motivation, and Difficulty × Aspiration × Information.

Design of Experiment

The task used in the experiment was dart throwing. In order to produce variation in performance the darts were thrown from distances of 8, 10, 12 and 14 ft. To control the strength of motivation, observations were made with the *S*s being rewarded in some cases and unrewarded in others. The rewards used were cigarettes. About half the observations were made during an acute cigarette shortage, and these *S*s were particularly glad to have the cigarettes. Two conditions of preliminary practice (designed to control the amount of information about the task which was available to the *S*s) were used, namely 10 and 50 trials of such practice. Two kinds of aspiration were required, namely a statement of hope and a statement of expectation. The experiment was cast according to the principles of factorial design. Hence, with the four distances, two conditions of reward, two conditions of preliminary practice, and two types of aspiration, it is clear that there were altogether 32 combinations of experimental conditions. A sample combination from the set of 32 was the requirement that an *S* throw darts from a distance of eight ft., under the conditions that she not receive a reward, that she have 10 preliminary practice trials, and at stated intervals make statements as to what she expected her final score to be. During the year 1944, 32 young women registered as students in the College for Women of the University of Pennsylvania were assigned randomly to the 32 conditions of the experiment, and a set of data was assembled from their work. This experimental work was done by A. S. During the year 1945 a second set of 32 young women from the same institution were assigned randomly to the experimental conditions and were studied by J. T. The experimental results therefore consist of a replicated set of observations made under each of 32 experimental conditions.

Details of the experimental procedure follow:

1. The dart board was mounted on a target four ft. square. The board was 15½ in. in diameter. On it were painted 10 concentric circles ¾ in. apart, numbered 10 to 1 from the center outward. The center of the board was 5 ft. 6 in. from the floor.

2. In each case the routine for an *S* was: a, preliminary instructions about the dart throwing requirement of the experiment, b, preliminary practice, c, instructions as to aspiration and first statement of aspiration, d, five throws at the target alternating with a statement of aspiration until 25 throws were completed. The foregoing procedure produced statements of aspiration after trials 5, 10, 15 and 20, in addition to the initial statement.

3. The *S*s who worked for a reward were given a cigarette for each 25 points accumulated at the end of the game. Up to 10 points could be made at a single throw; with 25 throws the maximum score was 250 and the maximum number of cigarettes to be won was 10. If the *S* was

working for a reward, appropriate information was given her in the preliminary instructions. The cigarettes were paid at the conclusion of the game.

4. The instructions on the level of aspiration for those *S*s who were requested to give realistic aspirations (expectations) were:

> Tell me what score you expect to make as a result of the 25 throws. By expect I do not mean the *least* you think you will make, nor the *most* that you might make if you were lucky, but rather I mean a *cool intellectual appraisal* resulting in the score which you think you will really achieve after the 25 throws.

The instruction on the level of aspiration for those *S*s who were requested to give irrealistic aspirations (hopes) were:

> Tell me what score you hope you will make as a result of the 25 throws. By hope I do not mean the *least* that you think you will make, nor what you think you *really will make*, but rather what you think you might make if you were lucky—or in particularly good form.

5. During the preliminary practice, as well as during the game, the *S* was required to remove the dart from the board after each throw, simultaneously calling the score aloud.

The data of the experiment consisted in a set of 64 game scores, a set of 64 initial aspirations, and sets of 64 aspirations taken after trials 5, 10, 15 and 20. Of the asiprations only the initial ones and those taken after trial 20 were analyzed.

TABLE I

MEAN GAME SCORE AS OBTAINED BY EACH OF THE TWO EXPERIMENTERS FOR ALL FRACTIONS OF THE DATA

Variable	Experimenter I	Experimenter II
Distance		
8 ft.	123.9	139.6
10 ft.	110.3	74.3
12 ft.	61.4	68.8
14 ft.	66.5	50.9
Preliminary Practice		
10 trials	82.6	78.1
50 trials	98.5	88.7
Incentive Conditions		
Prize	95.9	86.1
No prize	85.1	80.7
Accompanying Aspiration		
Hope	96.3	82.8
Expectation	84.8	83.9

RESULTS

I. Analysis of game scores

By a game score is denoted the accumulated score made by an *S* at the conclusion of 25 consecutive throws at the target. Two questions must be answered concerning these scores:

1. Were the effects of distance, prize and preliminary practice evident in these scores, and if so with what statistical reliability?

2. Were the results consistent from *E* to *E*?

An answer to these question is given in Table I, which gives mean game scores obtained by each of the two *E*s for the data fractionated for all variables. From the table it is evident that both *E*s found, in

general, that the mean game score decreased uniformly with increase of distance. In both instances the effect of preliminary practice was to increase the game score materially. In both instances the *S*s who worked for a prize made somewhat higher game scores. The *E*s differed only in what their data revealed about the differential effect of the aspiration upon the game score.

To determine the statistical realiability of the differences due to experimental variables disclosed in Table I, an analysis of variance was conducted with the error defined by the 32 differences between game scores from *S*s treated alike (except for the difference in the person of the experimenter).[1] The probabilities that the differences in Table I would occur in random sampling appear in the second column of Table II, which tabulates such probabilities from the

TABLE II

PROBABILITIES FROM ANALYSES OF VARIANCE OF GAME SCORES AND ASPIRATIONS

Source	Game Scores	Initial Aspiration	Final Aspiration
Distance	<.01	>.05	<.01
Prize	>.05	>.05	>.05
Practice	>.05	.05–.01	>.05
Aspiration	>.05	<.01	<.01
Distance×Prize	>.05	>.05	>.05
Distance×Practice	>.05	>.05	>.05
Distance×Aspiration	>.05	>.05	>.05
Prize×Practice	>.05	>.05	>.05
Prize×Aspiration	>.05	>.05	>.05
Practice×Aspiration	>.05	>.05	>.05
Distance×Prize×Practice	>.05	.01–.05	.01–.05
Distance×Practice×Aspiration	>.05	<.01	.01–.05
Distance×Prize×Aspiration	>.05	>.05	>.05
Prize×Practice×Aspiration	>.05	>.05	>.05
Distance×Prize×Practice×Aspiration	>.05	>.05	>.05

analyses of variance of initial and final aspiration as well as of game scores. The table indicates clearly that a significant change took place in the game scores as a result of change in the distance of the target. In this respect therefore, the experiment was capable of offering a basis for the distinction between the aspirations in the extent to which they were dependent upon performance. However, it is also apparent that the effect of preliminary practice and of motivation, while consistent from *E* to *E*, as well as in the direction of the differences produced, was not sufficiently evident in game scores to reach an acceptable criterion of statistical reliability. Finally, it

[1] In order to be certain that constant differences between the two *E*s in the scores and aspirations yielded by their *S*s did not vitiate this measure as a measure of error, preliminary tests were made which indicated lack of significance in the differences between the *E*s in these respects.

may also be noted from Table II that all interactions, double, triple and quadruple, are statistically without significance. This is important, since it constitutes evidence that the presence of interactions in the analyses of variance of the aspirations is not merely a reflection of similar interactions in game scores.

II. Analysis of initial aspirations

By an initial aspiration is denoted a statement of hope or expectation (depending upon what was required by the instructions), which was made following the preliminary practice and prior to the beginning of the game proper. Table III gives mean initial aspirations as obtained by each of the two *E*s for various fractions of the data. The table discloses the fact that there was only a slight systematic tend-

TABLE III

MEAN INITIAL ASPIRATION AS OBTAINED BY EACH OF THE TWO EXPERIMENTERS FOR ALL FRACTIONS OF THE DATA

Variable	Experimenter I	Experimenter II
Distance		
8 ft.	153.3	126.9
10 ft.	121.9	97.5
12 ft.	93.8	130.6
14 ft.	106.9	98.0
Preliminary Practice		
10 trials	106.3	99.0
50 trials	131.6	127.5
Incentive Conditions		
Prize	111.3	115.9
No prize	126.6	110.6
Aspiration		
Hope	147.8	133.1
Expectation	90.1	93.4

ency for the aspirations to diminish with increase of distance (in which respect the results from the two *E*s are not consistent), that the effect of preliminary practice was to increase the aspiration materially (a fact which appears in the data of both *E*s), that the effect of the prize upon the aspirations was inconclusive, and that a sharp distinction is made between reported values depending upon whether a hope or an expectation is required by the instructions.

From Table II it is apparent that the differences in the initial aspirations due to practice and to instructions possess statistical reliability. The difference due to instructions is, of course, to be expected, in view of work previously cited.

We may note immediately that the preliminary practice, while it did not produce significant differences in game scores, produced significant differences in aspirations. Such a fact suggests, of course, that the consequences of external controls on performance may well

be sought in other places besides the quality or the quantity of the performance which is subject to control. But we must also consider the significance of the increased value of the aspirations[2] with relatively extensive preliminary practice from the point of view of their realism. From Table II it is apparent that no significant interaction exists in the initial aspirations between Practice and Aspiration, from which it is clear that both hopes and expectations increased with practice. However, the two types of aspiration did not increase with practice in complete uniformity. The data show mean hope with 10 trials preliminary practice as 117.5, and mean hope with 50 trials preliminary practice as 16.3, mean expectation with 10 trials preliminary practice as 87.5, and mean expectation with 50 trials preliminary practice as 95.5. From these values it is clear that the effect of the preliminary practice was to increase hopes by 46.5 and to increase expectations by 8.0, i.e., the hopes were increased materially relative to the expectations.

The analogous values from the game scores are mean game score by *Ss* who yielded hopes with 10 trials preliminary practice 78.9, mean game score by *Ss* who yielded hopes with 50 trials preliminary practice 100.1, mean game score by *Ss* who yielded expectations with 10 trials preliminary practice 81.9, and mean game score by *Ss* who yielded expectations with 50 trials preliminary practice 87.1. It is apparent from these figures that the effect of the practice was to increase the game scores of those who 'hoped' by 21.2 and of those who 'expected' by 5.2. From the foregoing detailed analysis of the general increase in the aspirations resulting from the preliminary practice we must conclude that to some extent it is a reflection of a difference in game scores, which suggests that to some extent, at least, the increase must be regarded as reflecting an increasing realism in the aspirations.

The theoretical questions discussed in the introduction depend for their determination upon certain of the triple interactions in the initial aspiration, of which the relevant one, where practice is concerned, is that determined by Distance, Aspiration and Practice. This interaction raises the question as to whether hopes and expectations show the same differences in their dependence upon distance when practice is limited as they show when practice is extensive. That a significant triple interaction exists in this case can be seen in Table II where the corresponding P is less than .01. The nature of the triple interaction is seen in Fig. 1. The figure consists in four separate graphs. Each graph gives mean aspiration, either hope or expectation, as the case may be, at each of the four distances, from the data of *Ss* who were alike in respect to preliminary practice. The

[2] It should be understood that the mean value reported in Table III is a mean value over both expectations and hopes.

two practice conditions are denoted as P-10 and P-50. With each aspiration curve appears a curve (broken line) based upon the game scores of the Ss whose data contributed the aspiration curve. From the figure two conclusions may be formed, one having to do with the

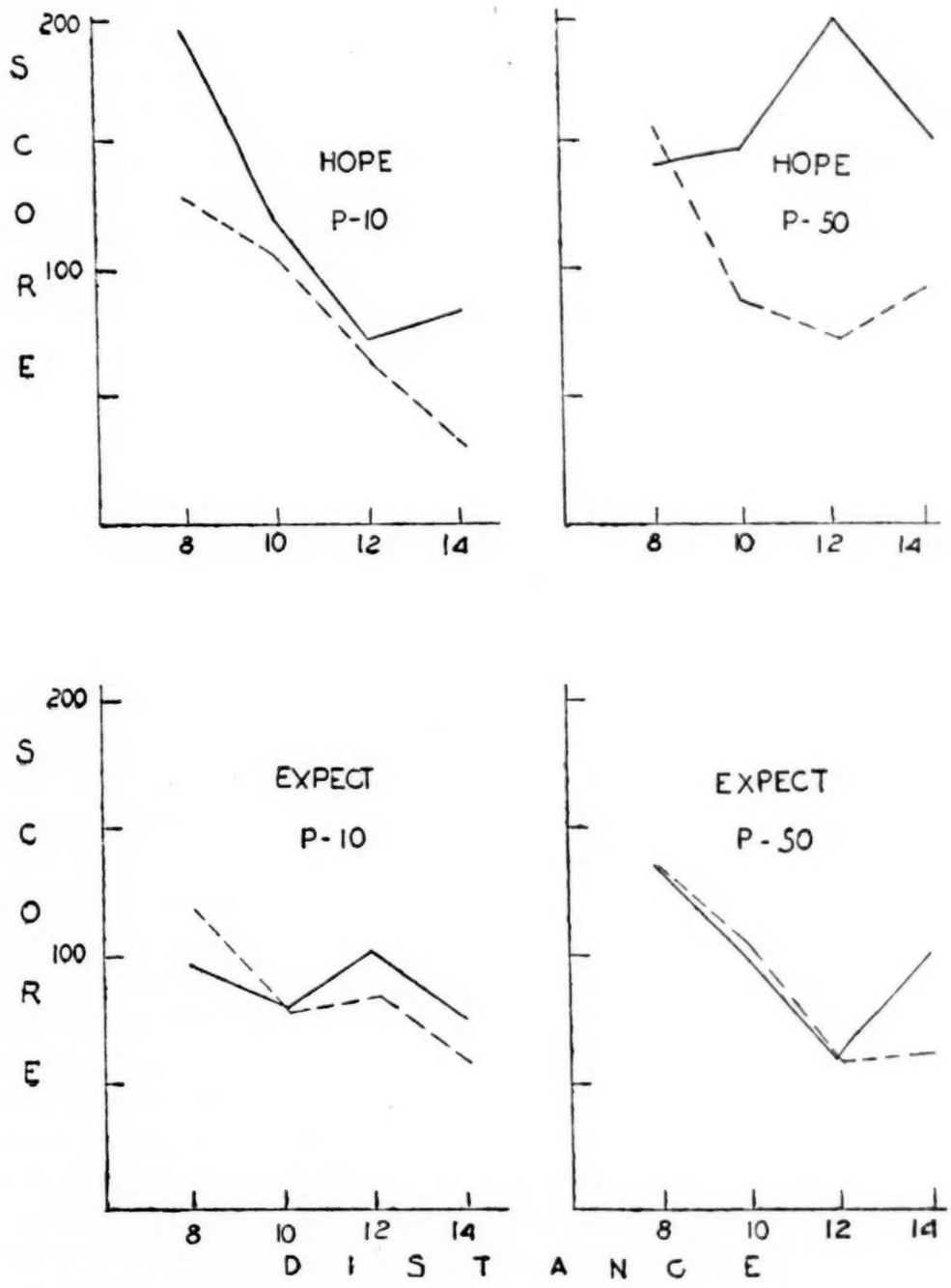

FIG. 1. Mean aspiration (———) and mean game score (– – –) as functions of distance and preliminary practice

properties of the significant triple interaction, and the other having to do with the extent to which this interaction is a consequence of a fortuitous variation in the game scores.

The figure is equally clear on both matters. In three cases

aspiration is clearly a decreasing function of distance; in the case of the hope aspiration, however, when 50 trials of preliminary practice are given (upper right figure), no such dependence appears. It is this fact which produces the significant triple interaction. Without exception, on the other hand, game score is a decreasing function of distance. For this reason we may conclude that under the circumstances where hopes are given and preliminary practice is sufficiently extensive, the reported hope will be relatively independent of the game score,[3] but where hopes are given and practice is limited the reported hopes will be dependent upon game scores.

Turning to the effect of increasing the motivation upon the aspirations, we may note from Tables II and III that prizes as a primary source did not produce significant effects upon the aspirations and that the interaction between prize and aspiration is insignificant. Furthermore, the triple interaction Distance × Prize × Aspiration is insignificant, from which we can conclude that on the average the double interaction Distance × Aspiration is of the same character no matter whether the *S*s are rewarded or not rewarded. In view, however, of the comparative failure of the reward to excite improvement in game score, we must regard this outcome as inconclusive for the hypothesis under test.

We should note, however, that Table II indicates the presence of a significant triple interaction of Distance, Prize and Practice in the initial aspirations with no such result in the same scores. While data will not be presented to illustrate the nature of this interaction, examination of the results indicates that it is a consequence of the fact that the double interaction Distance × Practice is significant for *S*s who were not rewarded and is not significant for *S*s who were rewarded. In the case of the *S*s who were rewarded, the mean aspirations of both practice groups decreased as distance increased. On the other hand in the case of the *S*s who were not rewarded, the mean aspirations *decreased* sharply if they were not practiced and *increased* sharply if they were practiced. Since no such interaction appears in the game scores, the interaction in the aspirations suggests that the *S*s who were motivated responded to the distance in much the same way no matter what the condition of practice, while the *S*s who were not motivated responded quite unrealistically if they were practiced. The meaning of this fact is obscure.

III. Analysis of final aspirations

By a final aspiration is denoted a statement of hope or expectation (depending upon what was required by the instructions) which was

[3] If the Hope P-50 figure is examined carefully, it will be seen that it offers some support for the view that the hopes tended to *increase* as the difficulty increased. Whether or not this view is correct we prefer to leave to be decided by later research.

made following the twentieth trial. At this point 80 percent of the game had been played and the final game scores, accordingly, were largely determined. For this reason one would expect these aspirations to be anchored to the game scores. Table IV gives mean final aspirations as obtained by each of the two *E*s for various fractions of the data. This table, studied in conjunction with Table I, which gives game scores, shows the influence of the game scores upon the aspirations. The aspirations decrease with distance, increase with practice and incentive, and are sharply defined by the instructions. Turning to Table II it is evident that a significant triple interaction is seen between Distance, Practice and Aspiration. While data will not be presented to illustrate the fact, examination of the data indicates that the effects apparent in Fig. 1 are present in diminished form

TABLE IV

MEAN FINAL ASPIRATION AS OBTAINED BY EACH OF THE TWO EXPERIMENTERS FOR ALL FRACTIONS OF THE DATA

Variable	Experimenter I	Experimenter II
Distance		
8 ft.	135.6	148.5
10 ft.	112.5	70.8
12 ft.	61.3	73.5
14 ft.	89.4	70.4
Preliminary Practice		
10 trials	96.6	82.8
50 trials	102.8	98.8
Incentive Conditions		
Prize	97.8	84.8
No prize	101.6	97.6
Aspiration		
Hope	116.9	101.5
Expectation	82.5	80.0

in the final aspirations, i.e., as distance increases with 50 trials of preliminary practice, expectations diminish uniformly, while hopes are independent of distance. On the other hand, with 10 trials of preliminary practice both hopes and expectations diminish uniformly with increasing distance.

The failure of the prize to be reflected directly in the aspirations is confirmed from the *P*-values of Table II. However, this table shows a significant interaction between Distance, Prize and Practice for final aspirations, as well as initial interactions. This interaction appears from the data to be of the same general character as that observed in the data of the initial aspiration.

DISCUSSION

The results of the present experiment, while decisive on the question as to the effects of prior information, are indecisive on the

manner in which motivation affects the realism of statements of aspiration. While it appears certain that the cigarettes failed to excite a difference in the motivation of the two groups, we cannot conclude that the experimental sample as a whole was unmotivated. The game situation itself made an appeal to the *S*s. On the other hand, we cannot conclude that they were well motivated. This lack of decisiveness is unfortunate in view of one interesting outcome of the experiment, namely the fact that under both conditions of motivation, hopes as well as expectations were functionally dependent upon game score, i.e., were both realistic. The determination of the conditions under which stated hopes and expectations were both *irrealistic* impresses us as an important matter, in view of its close relationship to the practical problems of personal adjustment. For this reason, if for no other, additional research on the question is justified.

Conclusions

1. If *S*s are relatively unpracticed in dart throwing, both realistic and irrealistic aspirations stated prior to the playing of a dart game will be dependent upon the game score and hence upon the difficulty of the task.

2. If the *S*s are given 50 trials of preliminary practice, realistic aspirations stated prior to the playing of the game will be closely dependent upon the game score and hence upon the difficulty of the task. With such preliminary practice irrealistic aspirations will be detached from any anchoring to the game score.

3. There is sufficient evidence that hopes *increase* with increasing difficulty of the task to warrant further research in an effort to define the character of the changes in hopes to be expected when the difficulty of the task is varied.

4. After 80 percent of the game has been played, both hopes and expectations are strongly anchored to the game scores. Despite this fact, the hopes from the practiced group are still relatively independent of game scores, while the expectations of this group and both the hopes and expectations of the relatively unpracticed group, are dependent upon game scores.

(Manuscript received March 15, 1946)

References

1. Irwin, F. W. The realism of expectations. *Psychol. Rev.*, 1944, 51, 120–126.
2. Irwin, F. W., & Mintzer, M. G. Effect of differences in instructions and motivation upon measures of the level of aspiration. *Amer. J. Psychol.*, 1942, 55, 400–406.

EXPERIMENTAL STUDIES OF THE JUDGMENTAL THEORY OF FEELING. VII. THE INFLUENCE OF NONMANIPULATIVE RESPONSES

BY HENRY N. PETERS

The Psychological Clinic, University of Hawaii

AND

FERN TALMADGE RODGERS

The purpose of the present study was to test the findings of previous experiments when the response determinants are of a more symbolic type. In the previous studies in this series (**2, 3, 4**) the responses used to modify affective judgments have been manipulations either of the stimulus objects or of the names of the objects. In the case of Japanese words, the printed word was either pronounced or not pronounced; with colors, the stimulus objects were named or not named; and in the experiment with pictures, the positive response was defined as reaching, touching, and pulling. It is obvious that many of our attitudes of pleasantness and unpleasantness in everyday life, such as prejudices, artistic judgments, and moral values, are not formed by any such directly manipulative types of responses. It is possible, although not readily conceivable, that the experimentally observed enhancing effect of positive responses has hinged upon the overt use, or repetition, of the peripheral mechanism associated with the stimulus. In the present experiment the overt positive response was the same for all stimuli in the group to which the *S* was required to react positively. Should the judgmental theory be verified under this condition, the likely interpretation would be that the crucial determinants of P–U changes are central modifications of the sort that attitudes are usually conceived to be.

Attempt was made to verify the major hypothesis, that positive and negative response tendencies underlie affective judgments, by the two methods employed in the study reported in the fifth paper of this series. These were, first, by demonstrating a shift in pleasantness of objects after positive and negative reactions are learned to them; and, second, by showing that the original response tendencies, identified in the pre-learning affective judgments, conflict with those established in learning in the form of interference with, or reinforcement of, the responses required by the learning problem.

In the two previous reports it was found that the concreteness of the *S*'s observing set is an important condition of the enhancing effect of positive responses on affective judgments. The results of these experiments indicated that the positive and negative response effect occurred when the set was concrete, but did not occur to any considerable degree when the set was abstract. This was given further investigation in the present experiment.

In addition, the experiment was designed to test a conflict hypothesis of affection, defining conflict as the existence simultaneously of two opposed responses to the same object.

Procedure[1]

Nineteen men and 15 women, students in introductory psychology classes, volunteered as *S*s. They were naive to psychological experimentation and were ignorant of the nature of the problem. In ages they ranged from 18 to 21 years. Every *S* came individually for two sittings separated by a week interval. The first sitting was for the purpose of determining original (pre-learning) affective judgments of the stimulus objects and the method, or set, used in making judgments. The second consisted of a learning situation which developed positive or negative attitudes toward certain ones of the objects and a second recording of affective judgments of all stimulus objects.

The first experimental sitting included two preliminary steps: taking the Ishihara color perception test and rating the affective values of 15 color reproductions of famous paintings on a nine-category rating scale. The Ishihara test was used to exclude any color weak *S*s. The rating of the pictures was primarily for the purpose of practice in giving value judgments under experimental conditions. The *S* sat while giving his judgments and looked at a large drawing of an equal-interval scale, which was hung on the wall in front of him. He gave his judgments in terms of numbers. The drawing of the scale looked very much like the drawing shown in Fig. 1, with the numbers inside of the slots, between the vertical lines.

The stimulus objects used in the experiment proper were 31 Hering colors. Colored rectangles, $1\frac{1}{2} \times 2\frac{1}{2}$ in., were mounted in the center of four × six in. white cardboards. In choosing the 31 colors the only criterion used was readiness of identification and of distinguishability from the others. The pack of cards was carefully shuffled before they were given to the *S* for rating. The colors were rated on a nine-category scale. The scale was drawn in heavy black lines on brown wrapping paper which was tacked to the top of a long table. The *S* stood, holding the pack of colors, and sorted them into the compartments. Thus, affective judgments were made by overtly placing the stimulus object into a position on the sacle. Fig. 1 is a reproduction of the scale. The numbers and the letters, with the exception of the U, I, and P, were not drawn on the scale used in sorting. The printed instructions defined the steps between adjacent categories as equal. No attempt was made to influence the method of rating or the number of colors to be placed in various compartments.

Immediately after the rating of the 31 colors, the *E* asked the *S* the following questions and recorded his replies.

(1) Now, can you tell me what your basis of judgment was?
(2) What were you thinking about while looking at the colors?
(3) Was it easy or hard to decide whether you liked or disliked most of the colors?
(4) Did you feel that you were being forced into deciding your preferences, i.e., that your decisions were arbitrary; or did decisions come readily and decisively?
(5) Do you think that most of these colors you either liked or disliked before you ever came to this experiment?
(6) Are there colors which you always like or dislike very much?

---

[1] The procedure used in this experiment is described in greater detail in a master's thesis on file in the University of Missouri library, under the name of Fern T. Rodgers.

The first question was framed to get a spontaneous report of method of judging. The remaining questions were considered pertinent in classifying the *S*'s method as an abstract set or a concrete set approach. If necessary, the *S* was urged to be specific in his answer to the first question. In the case of the last question, record was made of any colors the *S* named as liked or disliked.

At the second experimental sitting, a week later, the *S* was placed in a learning situation which involved nine of the 31 colors. A special list of 10 pairs of colors was constructed for every *S*. The method of constructing the list is illustrated in Fig. 1 and in the following list of pairs of colors used in learning:

$$\overset{- +}{X\text{–}Y};\ \overset{+ -}{A\text{–}B};\ \overset{+ -}{A\text{–}E};\ \overset{+ -}{A\text{–}D};\ \overset{- +}{B\text{–}E};\ \overset{- +}{B\text{–}C};\ \overset{+ -}{C\text{–}E};\ \overset{+ -}{C\text{–}D};\ \overset{- +}{D\text{–}E};\ \overset{+ -}{Z\text{–}W}.$$

The plus and minus signs indicate the type of response which was required during learning. The letters symbolize colors. The plus-and-minus over *E* in Fig. 1 signifies that this color was to be reacted to positively half the time and negatively half the time. The letters in these figures merely symbolize colors which the *S* rated in certain categories. Thus, A and B were any two colors placed in the third compartment. In selecting the nine colors for any *S* the *E* tried to avoid including two colors which might be hard to distinguish when presented in the visual field at different times.

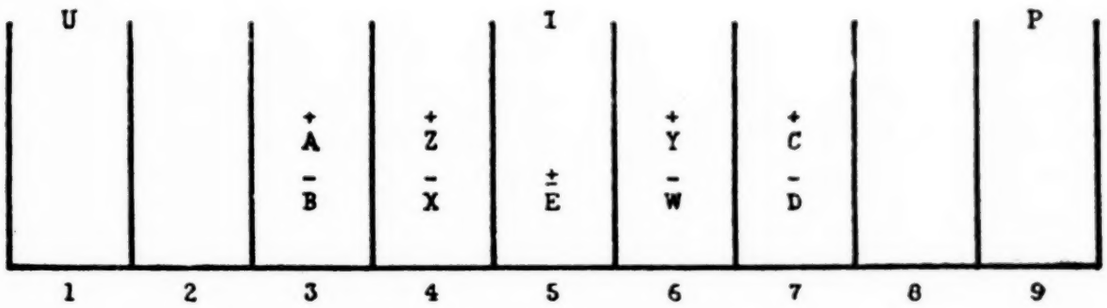

FIG. 1. The rating scale and location of colors used in the learning problem

After selecting the nine colors to be used, they were paired as indicated above. Since it was important to eliminate serial position effects for one phase of the problem to be investigated, three arrangements of the list were made up for the purpose of equalizing all place effects for colors A through E. In the three arrangements, pairs of which these colors were members were never placed in first or last position, but were placed in three different and widely separated serial positions. There were also right-left inversions of the two members of every pair. Pairs X–Y and Z–W occupied either the first or last position in all three arrangements. The arrangement shown in the text was not one of those used in learning.

The strips of colored paper used in making the learning list were $\frac{3}{8} \times \frac{7}{8}$ in., and were presented on a white background, the two members of a pair one-half in. apart. The pairs were exposed on an electrically operated Chicago Memory Drum. All but the window of the drum was concealed by a vertical gray screen, which also shielded the *E* from the *S*'s view. To the sides of the window in the drum, pasted on the screen, were two white cards, one with the word 'Right,' the other with 'Left,' printed on them in large black letters. These were to make the right and left sides easily identifiable and to remind the *S* of the manner in which he was to make his responses.

Printed instructions informed the *S* that pairs of colors would appear in the window, and that he was required to make a pointing response to one of the members of every pair by saying 'right' or 'left.' When he made a wrong response (picked the wrong color to indicate) or had not responded inside of a time interval of about three sec., a loud bell on the wall would clang. The *S* was informed that the learning process in the beginning would necessarily be trial-and-error in nature, and that the criterion of learning was five errorless trials, not necessarily in succession.

Every pair of colors remained in the window for six sec.; the bell rang, in case of error, after $4\frac{1}{2}$ sec. of exposure. The time interval between successive lists was 12 sec. The *E* recorded the correctness or incorrectness of response for every pair, for every trial. Although the particular colors used in the learning list varied from *S* to *S*, the arrangement of pairs on the drum was de-

rived from the *S*'s ratings according to a fixed pattern, which required the speaking out of 'right' and 'left' in a sequence which was constant for all *S*s. This facilitated the making and the handling of the records, since the same record blank could be used for all *S*s, and since the location of the response record for any particular pair was the same for all *S*s.

Results

I

The first step in the treatment of the results was the classification of the *S*s into a concrete set and an abstract set group, according to their method of perceiving the color stimuli. The answers of every *S* to the questions asked at the end of the first experimental sitting were typed on filing cards, every *S*'s answers on a separate card. The senior author read the answers on these cards several times and on three occasions, on three different days, rated every card as concrete, mixed, or abstract. Tabulation of the three ratings revealed a good deal of inconsistency. Only four *S*s were consistently judged to be concrete; 11 were judged abstract, and six mixed. Apparently division of a group of *S*s such as these into classes of abstract set and concrete set, from data such as the answers to our questions, is to some degree an arbitrary matter. It was decided to include in the concrete set group all *S*s who were judged on all three occasions to be either concrete or mixed and *S*s who were judged twice mixed and once abstract. The remaining *S*s, who were judged either abstract all three times or abstract twice and mixed once, were labeled the abstract set group. The concrete set group, on this basis, numbered 21 *S*s; the abstract set group, 13 *S*s.

Seven students in one of the senior author's advanced classes were used to estimate the objectivity of this grouping. Four class periods, of 50 min. each, were turned over to discussion of all phases of the abstractness of observing attitudes, how colors could be observed in an abstract and in a concrete manner, and the importance of abstraction to psychology and to life in general. Of course the cards were not read during this discussion. It was remarkable how difficult it is to make clear to students this idea of a continuum of abstractness, along which attitudes may be thought of as lying. Six of them seemed to grasp the idea finally; in one student's case, the writer was never certain that the idea had been understood. At the end of the discussion they came individually and sorted the 34 protocols, as the senior author had done, into concrete, mixed, and abstract categories. In averaging their ratings, a concrete judgment was given the weight of one, a mixed rating one-half, and an abstract rating zero. Thus, a total score was obtained for each of the 34 *S*s.

When these total scores were averaged, the means (and sigmas) of the previously described concrete and abstract set groups were,

respectively, 2.40 ± 0.52 and 0.46 ± 0.34; the medians, 1.4 and 0.0. Of the 13 abstract set *Ss*, 12 received no concrete ratings, three received only one (out of a possible seven) mixed ratings. The concrete set *Ss* were less consistently rated. However, only six received no concrete ratings; and only two of these were judged consistently to be abstract.

While the classification of the *Ss* into abstract set and concrete set must certainly be considered relative, it obviously has some objectivity to it. There probably never will be an absolute method of classification of abstractness of attitude. In the treatment of the results of this experiment, the classification based on the senior author's judgments was used.

II

Most of the *Ss* found the learning problem a relatively easy one. For the total of 34 *Ss*, the mean trials and errors, with sigmas, were 18.5 ± 1.2 and 38.6 ± 3.8.

In considering the influence of consistently positive or negative responses on affection, the changes in rating for all of the colors except E could be used. Table I presents, for the concrete and abstract

TABLE I

NUMBER OF SUBJECTS RATING THE COLORS IN DIFFERENT CATEGORIES AFTER LEARNING, WITH THE RATIO OF THE NUMBER OF TIMES THE COLOR INCREASED TO THE NUMBER OF TIMES IT DECREASED IN AFFECTIVE VALUE

(1) *Concrete set group*

Rating	A	B	C	D	Z	X	Y	W	E
1:	2	2	0	0	0	3	0	1	0
2:	2	7	0	0	2	4	0	0	0
3:	4	8	1	1	2	3	1	2	4
4:	6	3	1	1	2	4	1	1	6
5:	5	0	4	5	6	3	3	4	3
6:	0	1	1	7	6	3	7	8	4
7:	2	0	7	3	2	1	2	3	2
8:	0	0	5	2	1	0	3	1	1
9:	0	0	2	2	0	0	4	1	1
I/D	3.2	0.4	1.0	0.3	3.8	0.7	1.8	0.6	0.8

(2) *Abstract set group*

1:	0	2	0	0	0	1	0	0	1
2:	1	2	0	0	1	3	1	0	2
3:	5	3	0	0	2	3	1	0	0
4:	2	0	1	3	5	2	1	2	1
5:	3	4	1	2	2	1	1	4	4
6:	0	1	3	0	2	2	2	4	1
7:	2	0	3	4	0	1	4	1	4
8:	0	1	4	4	0	0	2	2	0
9:	0	0	1	0	1	0	1	0	0
I/D	7.0	1.5	1.0	0.8	1.7	0.6	1.8	0.5	1.2

groups separately, the numbers of times each of the colors was given every possible rating after learning. The italicized number in each column shows the number of *S*s who did not change their rating. The two members of adjacent pairs of colors in this table (A and B, C and D, etc.) received the same original ratings, but the first member was positive during learning, the second negative. The last row in both parts of the table shows the ratio of number of increases in rating to number of decreases. When this ratio is greater than one, more *S*s increased the rating than decreased it. Examination of both rows shows that the ratio alternately increases and decreases, as one would expect on the basis of the hypothesis that positive reaction enhances pleasantness and negative reaction depresses pleasantness. Both groups show the effect, although it is more variable in the abstract set group.

It is noteworthy that the effect of learning is about the same for colors A and C as for Z and Y, although the former received in learning many more of the correct positive responses. A and C appeared three times in every list presentation, whereas Z and Y appeared only once. The mean positive response indexes (number of responses divided by number of trials) for A and C were, respectively, 2.35 ± 0.07 and 2.51 ± 0.06; for Z and Y, 0.80 ± 0.03 and 0.87 ± 0.02.

TABLE II

MEAN DIFFERENCE BETWEEN THE SHIFTS IN SCALE INTERVALS FOR POSITIVE AND NEGATIVE COLORS

	N	*M*	σ_m	*t*	*P*
Concrete Set.....	21	4.05	0.86	4.72	<.01
Abstract Set.....	13	2.00	0.92	2.17	>.05

Table II presents the means for a score which combines the *S*'s rating changes for all of the critical colors. For each *S* the algebraic total of scale steps changed by the positive colors (A, C, Z, Y) was determined; the algebraic total of changes for the negative colors (B, D, X, W) was found; and the latter was subtracted from the former. This gives a measure of the extent to which the *S* showed the expected shift in affective values. These scores were used in calculating the means of Table II. The expected trend is greater and more reliable in the concrete set group than in the abstract set group.

Several different methods of analyzing the data of this experiment were tried, and results the same as those shown in Table II were found in all cases. Although the concrete group yields a large and reliable effect and the abstract group yields a small and unreliable shift, the difference between the two groups is not reliable.

III

The second phase of the central problem of this study concerns the relationship between the relative difficulty of learning and the original affective values of the colors. From the judgmental theory we have deduced: (1) that it should be easier to learn a positive response to a pleasant stimulus than a negative response, and easier to learn a negative response to an unpleasant stimulus than a positive response. And (2) that *S*s of concrete set should show this interference, or reinforcement, effect (or affective value on learning) more than abstract set *S*s.

The reasoning behind these deductions is fully developed in the fifth paper of this series (3). It may be briefly summarized as follows. P and U judgments indicate the presence of positive and negative reaction tendencies (attitudes). Learning is the establishment of positive and negative reaction tendencies. Hence, the latter should be reinforced by corresponding reaction tendencies, as indicated by previous affective judgments, and should be interfered with by opposed reaction tendencies. That these effects should be greater in concrete set *S*s than in abstract set *S*s is deduced from the postulate that abstract reaction tendencies are more rigid, more highly learned, than are concrete reaction tendencies. This deduction is supported by the demonstration in a study of retroactive inhibition that well learned reactions are less interfered with by subsequently learned reactions than are weakly learned ones (1).

TABLE III

MEAN ERROR-INDEXES FOR THE CRITICAL COLORS

	A	B	C	D	E
Concrete Set	.70±.09	.45±.08	.50±.07	.69±.08	1.00±.08
Abstract Set	.49±.08	.59±.13	.50±.09	.60±.11	.90±.09

The critical colors for investigating this relationship are A, B, C, and D, which had their serial positions rotated during learning in such a way as to eliminate place effects on difficulty of learning. The measure of difficulty used was relative numbers of errors. Table III presents the mean error-indexes (E/T) for the critical colors. The actual number of errors any *S* made on a color was divided by his number of trials to learn the whole problem, giving the E/T for that color. The means are derived from these relative scores, which tend to reduce the effect of the variable speed of learning the whole problem.

As the reader can verify by referring back to Fig. 1 and the procedure, colors A and B were U colors of the same original rating; to

the first a positive response was learned, to the second a negative response was learned. In the same way, C and D were P colors of the same original rating: to the first a positive response was learned, to the second a negative response was learned. Hence, it was expected that there would be greater numbers of errors on A than on B; and more on D than on C. This is what happened with the concrete set *Ss*. The differences are in the direction expected and are fairly significant in reliability. The CR for the difference between the means for A and B is 2.03; for C and D, CR = 1.74. The first difference is opposite to expectation for the abstract set *Ss*, the other is small and unreliable.

A second method of approaching this problem is by comparison of mean errors for *pairs* of colors. For six of the pairs which involved critical colors, the requirement of avoiding one and pointing to the other was either in accord with the tendencies inferred from the P–U

TABLE IV

MEAN ERROR-INDEXES FOR VARIOUS PAIRS OF COLORS

Set	(1) Means for Pairs					
	A–D	B–C	A–E	B–E	D–E	C–E
Concrete.....	.220±.033	.103±.033	.296±.037	.166±.038	.297±.043	.229±.036
Abstract.....	.167±.032	.177±.045	.176±.045	.265±.063	.279±.062	.175±.043
	(2) Mean Differences					
	(A–D) – (B–C)		(A–E) – (B–E)		(D–E) – (C–E)	
Concrete.....	+.118±.036		+.130±.052		+.069±.046	
Abstract.....	−.010±.058		−.089±.087		+.104±.081	

judgments or directly opposed to them. In the case of pairs A–D, A–E, and D–E the *S* was required to go against his original tendencies; while in the case of pairs B–C, B–E, and C–E he was required to make responses which agreed with his original tendencies.

The first part of Table IV presents mean E/T's for these six pairs of colors. Those pairs between which a significant comparison can be made are presented adjacently. Thus, all factors, except agreement in reaction tendencies, are the same for A–D and B–C; this is also true for A–E and B–E, and for D–E and C–E. The concrete group means are in all three cases higher for the first pair, which is opposed to original tendencies in learning requirement, showing a tendency toward greater difficulty in learning. On the other hand, with the abstract group the difference between the means is in two cases opposed to expectation.

The second part of this table shows the mean differences for the adjacent pairs of colors. These means were derived from the differences in E/T for every *S* in the group, calculated individually. Hence the pertinent CR's can be gotten by simply dividing the sigma into the mean. The first of the concrete mean differences, that between A–D and B–C, is highly reliable; the next is significant; the third has a CR of 1.5.

IV

In the learning problem color E, which was originally judged to be indifferent in affective value, had to be responded to half the time positively and half the time negatively. The required response was consistent only in being positive when E was paired with two of the other colors and negative when it appeared with the other two.

A common definition of 'conflict' is that it consists of the simultaneous existence of two opposed reaction tendencies to the same stimulus objects. The idea of conflict between impulses or neural pathways, as a cause of unpleasant feeling tone, has appeared frequently in past theorizing about affection. In the present experiment the opposed reaction tendencies to E were planned as at least a beginning toward investigation of this hypothesis.

One would deduce from the conflict theory that E should decrease markedly in affective value. The tabulation of results in the last column of Table I shows clearly that this did not occur. The average shift, as measured by the number of increases over the number of decreases in scale position after learning, is in both groups of *S*s intermediate relative to the shifts in the other eight colors.[2]

In view of this result, either the conflict theory of P-U will have to be discarded, or it will have to specify a definition of conflict more inclusive than merely the simultaneous existence of two opposed reactions.

Summary

Although the conditions of this experiment and those of the investigation reported in the fifth paper of this series (3) differed widely, the results of the two are in fundamental agreement in verifying the judgmental theory. Both sets of results show an enhancing effect of learning a positive reaction and a depressing effect of learning a negative reaction on affective judgments. Both also show that the reaction tendencies, inferred from P-U judgments, which are origin-

[2] Positive and negative reactions were required to the same stimuli in the study reported in (3), although the two opposed reactions were not required an equal number of times. The results agree with the present ones in showing that this kind of 'conflict' apparently had no effect upon affective value.

ally attached to the stimulus objects either conflict with or reinforce the reaction tendencies required during learning.

The results of the two experiments are also in agreement in showing that the above two relationships, one of affective shift and the other of interference, are conditioned by abstractness of observing set. When the set is highly abstract, affective judgments are more fixed and unchangeable, at least as far as subsequently learned reaction tendencies are concerned, and the reaction tendencies inferred from the judgments do not seem to interfere with the subsequent learning of reaction tendencies. On the other hand, the two effects appear clearly in the performance of Ss who judge under a relatively concrete set.

The two experiments show that the frequency of positive responses is not *directly* related to the enhancement of affective value. Although the objects which increase in affective value apparently have to be reacted to positively during learning, the absolute or relative number of times the overt positive reaction takes place does not seem to be important.[3] The crucial factor seems to be that the object is merely identified as one to be reacted to positively. This favors the interpretation of the real causal agent in these experiments in terms of centrally localized, positive and negative *attitudes*. This interpretation is likewise favored by the fact that the enhancing effect of positive reactions occurs, as it did in the present experiment, when the overt reactions are purely symbolic, non-manipulative in nature, and are variable in form. Apparently the creation of positive (or negative) attitudes, defined as reaction tendencies, does not depend upon the exercise of a particular peripheral response mechanism.

The part of the present investigation which was directed at testing a conflict theory of affection failed to verify that hypothesis.

(Manuscript received March 20, 1946)

References

1. McGeoch, J. A. The influence of degree of learning upon retroactive inhibition. *Amer. J. Psychol.*, 1929, 41, 252–262.
2. Peters, H. N. Experimental studies of the judgmental theory of feeling: IV. Retention of the effects of learned reactions on affective judgments. *J. exp. Psychol.*, 1939, 24, 111–134.
3. Peters, H. N. Experimental studies of the judgmental theory of feeling: V. The influence of set upon the affective values of colors. *J. exp. Psychol.*, 1943, 33, 285–298.
4. Peters, H. N. Experimental studies of the judgmental theory of feeling: VI. Concrete versus abstract sets in the preference judgments of pictures. *J. exp. Psychol.*, 1943, 33, 487–499.

[3] Refer to (3), Table IX, p. 296.

THE MODIFICATION OF JUDGMENTS IN A GROUP SITUATION*

BY ROSALEA ANN SCHONBAR

Wheaton College

The phenomenon of conformity to social standards of behavior has been studied from many different points of view. Statistical devices for the determination of conformity behavior have been set forth by some observers (**2**), while others (**3, 5**) have emphasized the influence of various pressures which produce uniform social behavior. Other studies have investigated the formation of social norms, and conformity to them as they are developed (**8, 9, 10**).

While relevant to most of these, the present investigation stems from a slightly different approach: that of F. H. Allport (**1**), reported in 1924. It attempts to answer the following question: does the presence of a coacting group operate to cause the individual to give less extreme judgments than he would render when alone?

Allport had his *Ss* rate the pleasantness of a series of odors on a device similar to a graphic rating scale. He found that when this was done in a group, with each observer writing his judgments individually, and with no interaction among them, pleasant odors were judged less pleasant, and unpleasant odors less unpleasant, than when the observers had previously judged them in individual sessions. The range of judgment was thus narrowed in the social session, extreme judgments regressing toward the center of the scale, and intermediate judgments being affected little if at all.

Allport ascribed his results to an attitude of social conformity, the result of unconscious submission in the presence of a group. The range of judgment narrows, according to Allport, because of the conscious or unconscious thought that one is less likely to be deviant in the group situation if one's judgments are less extreme. Similar behavior may be cited in the social field in which individuals modify their behavior, at least temporarily, so that extremes are avoided, and possible differences between individuals thus narrowed. An instructor may conceivably be less inclined to give extremely high or extremely low grades on questions which he knows are to be read by other instructors; some people are loath to express an opinion on a

* The research for this paper was performed as part of a dissertation for the degree of Doctor of Philosophy at Columbia University under the direction of Professor Otto Klineberg.

controversial issue until they have some idea of the views of the persons with whom they are engaged in discussion. Gallup (6) points out that with poll questions of the multiple-choice type, respondents tend to avoid the extreme choices, selecting instead the intermediate alternatives.

Granting that this levelling type of behavior does sometimes occur, we may still ask if it is *typical* of such social group situations. Allport's results have not as yet been confirmed by other investigators. Farnsworth and Behner (4), using more refined controls, failed to find the clearcut trend predicted by Allport's results. In 1933, Perl (7) found that the previously rated funniest and least funny of a series of jokes were judged to be further apart in social than in individual sessions.

In view of these contradictory results, it seems wise to examine Allport's investigation with a view toward more rigid experimental control and variation. Two main modifications in procedure appear indicated:

1. All of Allport's *S*s judged first alone (A), and then in the group situation (T = together). Since the apparent modifications in judgment, when plotted, greatly resemble those of a regression or series effect due to temporal interval,[1] it would appear wise to divide the group in half, half judging first alone and then together, and the others in the reverse order. This procedure was adopted by Farnsworth and Behner, but in their presentation of data, the results of the two groups are combined, so that it is impossible to determine whether or not they behaved differently. However, their total results indicate that this procedure yields data which are different from Allport's. To isolate still further any possible temporal effects, a third group seems advisable, to judge the stimuli individually, with no opportunity for social intercourse among the observers.

2. As has already been noted, all of Allport's *S*s rendered their judgments in writing, thus precluding the possibility of any real social interaction, and heightening the possibility of the operation of factors other than social. Verbal judgments would seem to be necessary, to give at least a minimum of social 'togetherness' to the situation.

Both of these major modifications in procedure will be adopted in the present study.

[1] The so-called central tendency of judgment or series effect, first commented upon by Hollingworth, was described by him as follows: "In all estimates of stimuli belonging to a given range or group we tend to form our judgments around the median value of the series . . . toward this mean each judgment is shifted by virtue of a mental set corresponding to the particular range in question." Thus extreme judgments tend to regress toward the mean, while medium judgments are not affected.

Materials, Subjects, and Procedure

The choice of materials for this investigation was not an easy one. Several kinds of stimuli were tried and discarded. It was felt that materials which were not members of an objective series might be desirable, since they would allow the greatest possible range of freedom in individual judgment. Some form of esthetic or affective judgment seemed to be called for, if we were in any way to approximate Allport's original judgment situation. Preliminary investigation revealed that abstract esthetic stimuli were clumsy to work with, and provided very little interest value to motivate the observers.

The stimuli finally decided upon were a series of yearbook photographs of members of a senior class at Columbia College. These were of uniform size (2½ in. × 1½ in.), of uniform gloss, showing the head and shoulders of each young man; of the original group of 150 photographs, all those with mustaches, eyeglasses, or similar distinguishing features were omitted. From the remaining group, three *S*s cooperated in choosing a series of 10 which seemed to them to represent a series of pictures in scale from relatively attractive to relatively unattractive.

The *S*s were all female undergraduates, sophomores and juniors at Queens College, who were enrolled in courses in General or Child Psychology.

The instructions to the *S*s were as follows:

> We are trying to standardize a test of something we might call social esthetics. This part of the test deals with the judgment of physical attractiveness.
>
> You know that when you first meet or see a picture of a boy, you decide on the basis of your own standards how attractive you think he is . . . not how intelligent, or of what kind of character, but just whether or not you find his appearance to be attractive and would care to be seen with him. I have here ten pictures of boys, and I want you to judge them in terms of *your* standards of what constitutes an attractive boy . . . not to compare them with each other, but to judge each with reference to your own standard.
>
> I am going to ask you to assign a percentage in attractiveness to each one, in the same way that an Art instructor grades the efforts of his students. Each one is to be graded on his own merits, not on how he compares with his fellows. You will see each picture three times.

The pictures were then shown to the *S*s, a few trial judgments were taken in order to familiarize them with the kind of judgment they were to give, and then the pictures were presented, three times each, in random series. The *S* was given pencil and paper, and was instructed to give her judgment aloud, and then to write it down. This procedure was followed in both the alone and the together situations.

There were 26 *S*s in all, completely naive and inexperienced in psychological observation. Ten of these (Group AT) judged first alone and then in the group. Six (Group TA) judged first together and then alone. The group judgments were given in a group of 25 girls, comprising our 16 experimental *S*s and nine other girls who judged only at this time, and whose data are not part of the results to be presented. The pictures were passed around from individual to individual, one at a time, each girl judging aloud and then recording her judgment in writing. The order of judgment was varied by starting the passing around of the pictures in different parts of the classroom, and by having them circulated in different directions. It is not suggested that this is an ideal 'social' situation, but it at least provides for a minimum of social interaction in that each girl's judgments are rendered in the hearing of a coacting public of her contemporaries and fellow-students. Ten other girls (Group AA) make up our control group, giving two sets of individual judgments.

Twenty-four hours separated the first from the second judgments of all individuals. The *S*s were requested not to discuss what they had done with any other members of the class; inquiry later did not reveal that there had been any such discussion. The second judging period was explained in terms of obtaining data concerning the reliability of the 'test being standardized.' The *S*s were familiar with the concept of reliability, but were naive enough psychologically not to question the difference in conditions of the two judging periods.

Results

In order that our results might be directly comparable with Allport's, we adopted the same procedure for presentation and analysis

of our data as he did in the odor experiment. We averaged each *S*'s judgments of attractiveness for each picture under each condition. We then determined the rank order of preference for the pictures from these mean individual estimates; these individual rank order means were then combined, and the mean judged percent was found for each position in the rank order. These mean judgments were then plotted against rank order. Thus, the average at any point on any of the graphs which will be presented will cut across the actual stimuli to some extent, although there was a great deal of uniformity of judgment.

In addition, we determined whether or not the differences between the two sets of judgments for each of the three groups were significant. The *t*-test was used, and a *P*-value of .05 or better was accepted as our level of significance.

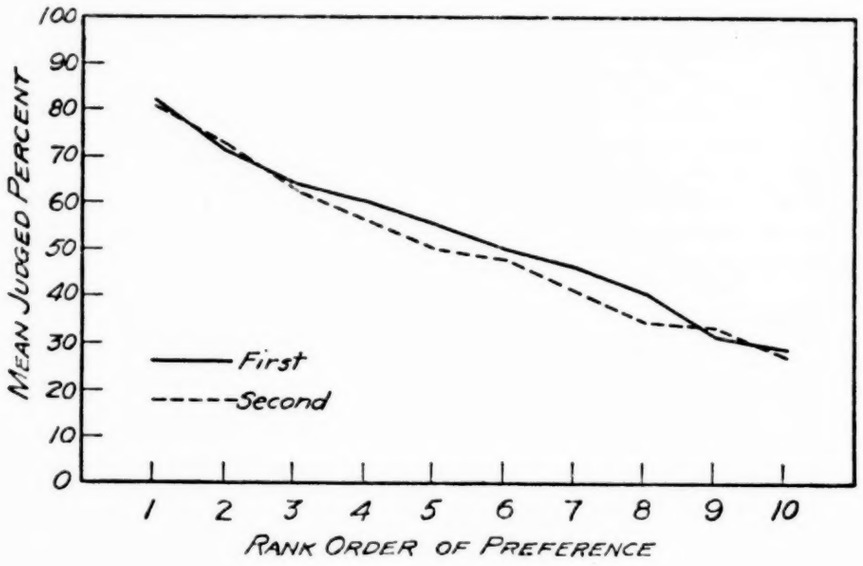

FIG. 1. Mean judged percents, first and second sessions: control group

Fig. 1 presents the data of Group AA, our control *S*s. These curves represent the mean judged percent of attractiveness for each of the 10 pictures in rank order of preference by *S*s judging them in two alone sessions. As we can see, eight of the 10 pictures were judged to be slightly less attractive in the second session. None of these differences, however, was significant according to our criterion. Thus, for all intents and purposes, judgments of the physical attractiveness of college boys from their pictures do not change appreciably after 24 hours. Nevertheless, although this may be statistically true, the insignificant change which does take place is almost consistently in one direction, namely, lower in the second session. This slight and insignificant depression of judgment in the

second judgment session will serve as a basis of comparison for our two experimental groups.

It will be noted that both the unidirectionality of the slight shift and the relatively greater change in the center of the scale eliminate the possibility of the operation of the central tendency of judgment. Thus, Allport's results must be attributable to something more than this temporal influence.

Turning now to Fig. 2, which presents the combined data for the alone and together sessions of our two experimental groups regardless of temporal order, we note that here, too, there is a marked tendency for the pictures to be judged less attractive, this time in the social as compared with the individual situation. Again, eight of the 10 stimuli followed this trend, the two exceptions being near the positive

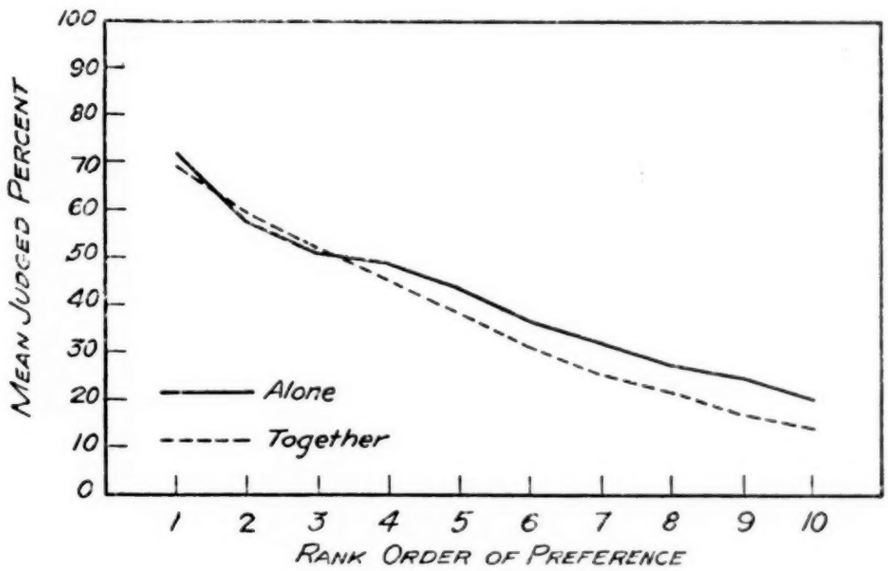

FIG. 2. Mean judged percents, alone and together: experimental group, combined

or attractive end of the scale. Again, none of the differences was significant at the .05 level.

We conclude therefore from the data thus far presented that judging in a group produces no changes which would not also occur simply by judging alone twice. Figs. 1 and 2 reveal a depression of judgment in the second and social situations respectively. In neither case is this a significant shift. In neither case is there a differential effect on direction of the change in judgment. Whatever change occurs is a unidirectional one in both instances. Furthermore, there is nothing in Fig. 2 (or in any of the individual curves which make up this figure) which resembles Allport's results.

It will be remembered that the data presented in Fig. 2 are combined data for two subgroups who differed in the temporal order of

the individual and social sessions. Separate data for each of these groups are presented in Figs. 3 and 4.

Fig. 3 presents the mean judgments of Group AT, who judged first alone and then together. Here again we note the depression of judgment in the social as compared with the individual session. For this group, however, the trend is absolutely consistent, and the differences between the points on the curves are all significant except that at the most favored position. In other words, this group, the order of whose judgments was like that of Allport's *Ss*, does not yield results which in any way resemble those obtained by him. There is here a consistent, significant, unidirectional shift toward lower group judgments rather than a drawing in from the extremes, and the change is at least as marked in the center of the scale as at the extremes.

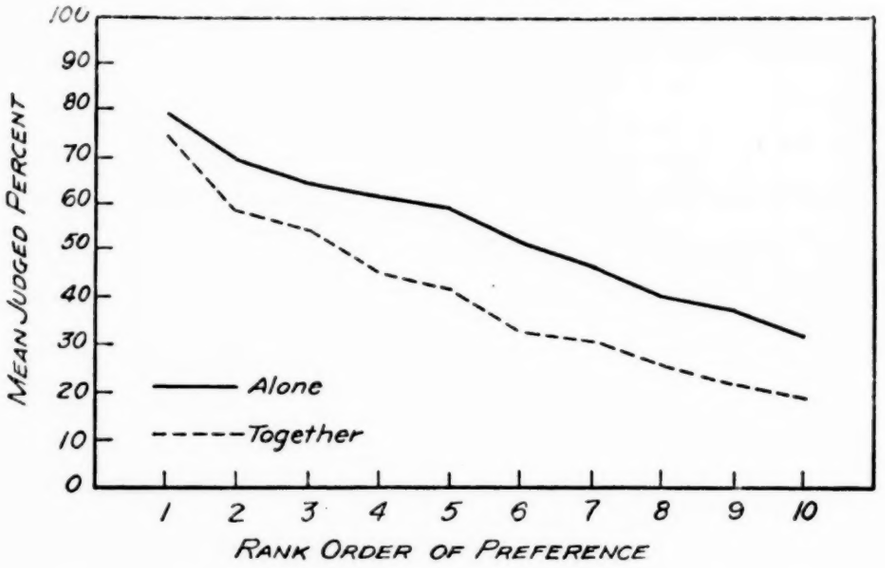

FIG. 3. Mean judged percents, alone and together: Group AT

There is also one important difference between the results of Group AT and those of our control group. Although the change in attitude toward the attractiveness of the boys is in the same direction for both groups, the trend for Group AT, unlike that for Group AA, is a statistically significant one. What difference is there in conditions which could account for this difference in results? The only factor which was not identical for the two groups was the social factor. In the group where this factor was present in the second session, the judgments in that session were reliably lower. This seems to indicate that the effect of judging in a group is that of causing the *Ss* to find the pictures less attractive. Since this effect was found not to be significant when we combined data for this group with those of the other experimental *Ss*, then it would appear that the trend becomes

significant only when the temporal and social 'forces' are not opposed, when, in short, the social session is also the second session. In this case, the two factors augment and reinforce each other, resulting in a significant depression of judgment, whereas either factor alone produces only an insignificant shift in the same direction.

This hypothesis is further borne out when we examine the results of the six *Ss* in Group TA, where the social and temporal tendencies are in opposition. These data are presented in Fig. 4. This group presents quite a different picture from Group AT. Here individuals find the pictures to be *more* attractive in the group situation. This is true along the entire scale save the most preferred position, where there is no difference. Only two of the differences are significant. If we disregard for a moment the social aspect of the situation, we

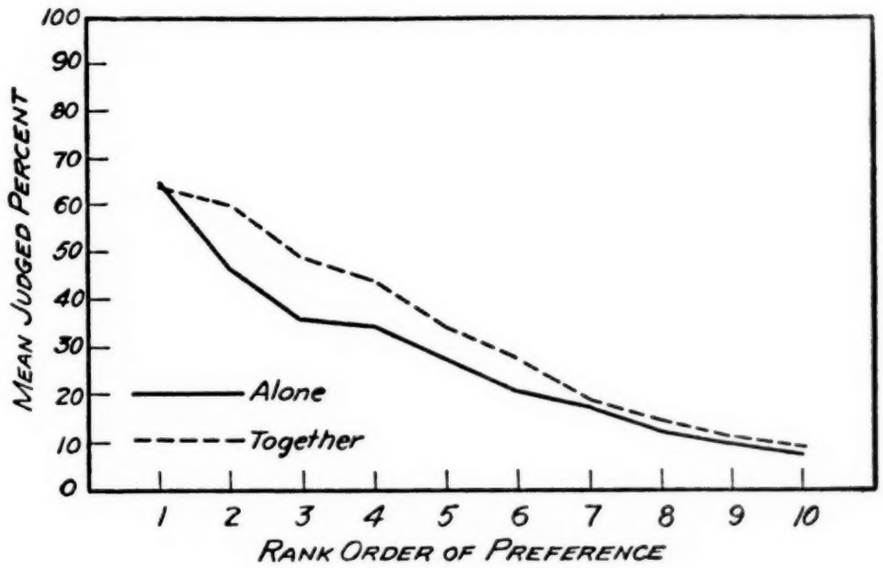

FIG. 4. Mean judged percents, alone and together: group TA

note that the results are very similar with respect to first and second judgments to those obtained from the control group. The judgments are higher in the first situation than in the second; this despite the fact that the first session was also the social session. Thus, we may say that when the social factor and the temporal factor are in opposition, judgments tend to be rendered in accord with the trend characteristic of the temporal influence rather than the social influence. This explains why the differences between the alone and together situations for the combined groups are not significant. It is true, however, that the whole curve for the first (social) session reveals lower judgments than any other first session.

Fig. 5 represents the final comparison which is to be made; this figure shows the combined data for our experimental *Ss* for first and

second sessions regardless of the position of the group session. As is to be expected, judgments are lower in the second than in the first situation. This is true for all stimuli, and the differences between sessions are significant for all save the most preferred picture. These findings corroborate the conclusions we have already drawn from the rest of the data. All of the previous findings emphasize this tendency toward depression of judgment in the second session. However, this tendency, operating alone in the control group was not significant. Here, on the other hand, part of the tendency (for part of the group, that is) is supplemented by a social influence operating in the same direction, thus causing the differences to become significant. The fact that where the social and the temporal factors are irreconcilable, the temporal influence is the dominant one and takes precedence over the other, also contributes to the significance of the difference here.

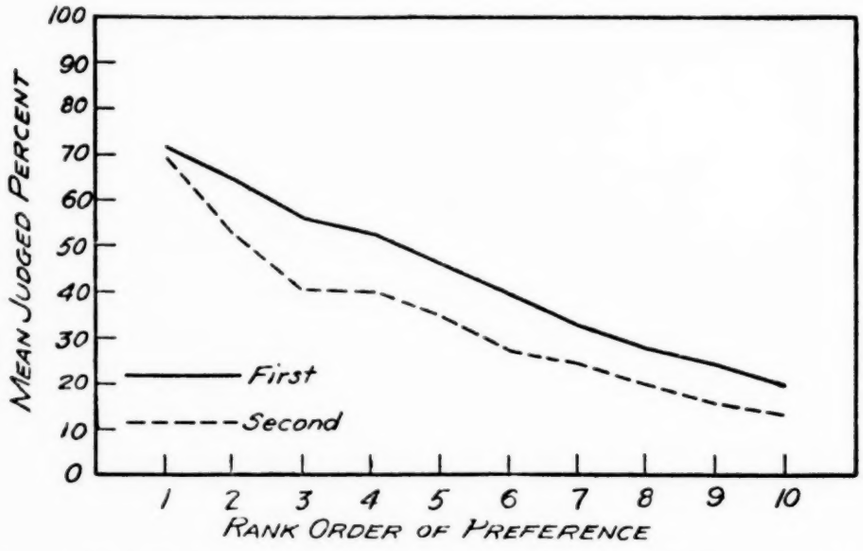

FIG. 5. Mean judged percent first and second experimental groups combined

With reference to the range of judgment, there appears to be no constriction in the social as compared with the individual judgments. Rather, there is a slight increase in range for Group AT, which is compensated for by an almost equivalent decrease in range for Group TA in the group situation, the two tendencies thus cancelling each other. This means that as we go from first to second judgment session, there is a consistent widening of range, whether that second session be social or individual. This widening of range is small and unreliable, and for the control group, although it lies in the same direction, it is small enough to be negligible. These figures are presented in Table I.

Summing up now the trends which have been noted in this investigation, we find that when we isolate the social and temporal

factors which might effect shifts in judgment, the temporal factors play the more dominant rôle. Judging in a group does not seem to be so important with respect to changes in judgment as does judging for a second time after a lapse of time. The evidence reveals that the effects of judging in a group are in the same direction as those of judging for the second time. In both instances there was a consistent but insignificant lowering of judgments. When the group judgment is also the second judgment, these two similar trends reinforce each other, and the shift becomes significant. When, however, the social judgment is the first session, the effect of judging in a group is almost completely lost, and the individual second judgments are lower than the social first judgments, indicating the dominance of the temporal over the social factor. With respect to constriction of the range,

TABLE I

RANGE OF JUDGMENT, FIRST AND SECOND SESSIONS

Group	First	Cond.	Second	Cond.
AT	47.2	A	55.8	T
TA	48.3	T	55.6	A
AA	53.1	A	54.1	A

which Allport found to result from social judging, no such effect was found; a slight broadening of the range occurred in the second session regardless of the position of the social situation.

One thing becomes immediately apparent: our results in no way resemble the results of Allport. Far from discovering that *less* extreme judgments are given in the social session, we find a unidirectional shift which results in far *more* extreme judgments being given at that end of the scale in which the direction or the shift lies, that is, at the low or unattractive end of the scale. At the preferred end, where the shift is in the direction predicted by Allport, we find the most limited change. With respect to the direction of the shift, our results are more similar to those of Farnsworth and Behner and of Perl, both of whose studies also reveal unidirectional changes in judgment in the social session. We cannot then ascribe the difference between our results and Allport's to the difference in stimuli, at least not wholly, since similar results have also been found with other stimulus materials, including one which was identical with the stimulus materials of one of Allport's experiments. A next step, however, would be the duplication of his stimulus series in combination with the more controlled procedure.

On the other hand, our stimulus material is unique in one respect: the materials to be judged are more salient to the judges than were

the others. This is the type of judgment which our Ss implicitly render practically every day. It is the type of judgment for which there are already rather strong social norms operating. For college girls, it may well be that the desire to conform to the group norm takes the form of 'playing down' one's estimates of the physical attractiveness of contemporaries of the opposite sex. It may also be that the individual feels that she is less likely to be deviant from group judgments if she becomes more critical of the pictures, thus making her judgments lower. Or it may simply be that she actually becomes more critical as a direct effect of the group situation. We are able to say this despite the seemingly opposite tendency of Group TA, since, when the first judgments are rendered in the group, the whole range of judgment becomes lower (as compared with Groups AA and AT), and the inter-session shifts therefore occur on a lower level. Whatever of the above social factors is responsible for the lower group judgments, the process is not a conscious one, since none of the Ss reports that she was aware of any differences between her individually and socially rendered judgments.

It is unfortunate that we are unable to analyze Farnsworth and Behner's results for their two experimental groups separately. It will be remembered that their experimental Ss were divided temporally in the same way as ours. Our results stress the importance of separating out any temporal factor in experiments of this sort before drawing conclusions as to possible social influence. Our results indicate that the temporal factor may completely overshadow whatever tendency the social factor may evoke. Thus, even though our control group reveals no more of a regression effect than did our experimental Ss, we must be doubly wary of accepting the trends reported by Allport as being attributable solely to the group judgment, especially since his Ss could not be directly influenced by other members of the group, but only by their own estimates of how the other members of the group were judging. It is an interesting detail that the only one of our 26 Ss whose results were similar to those of Allport's Ss was a member of Group AA, our control group which judged alone twice.

Summary and Conclusions

The purpose of this experiment was to ascertain whether rendering judgments within a group caused individuals to modify their judgments in the direction of becoming more moderate, of giving less extreme judgments. We wished therefore to see whether judging as a member of a coacting group exerted a differential effect upon favorable and unfavorable judgments, and also whether or not the

range of judgment was affected. In an attempt to isolate this social factor from all other factors, mainly from any effect which might obtain merely because the individual was rendering two sets of judgments, our *Ss* were divided into three groups: one which judged alone twice, one which judged alone and then together, and one which judged together and then alone. The material to be judged was a series of photographs of members of a senior class at Columbia College; the *Ss* were female undergraduates at Queens College. Each *S* rendered three judgments of the physical attractiveness of 10 photographed boys at each session; these judgments were then averaged for each stimulus, and arranged in the rank order of preference for each *S*. A mean was then found for each group for each position of the rank order, and this mean was then plotted against rank order. The significance of the differences between the means for each session was also determined. From the data presented, we may draw the following conclusions:

1. There is no differential effect evoked by judging in a group which causes extreme judgments to be avoided. Rather, the social factor causes a slight and insignificant depression in judgment, the pictures being regarded as less attractive in the group session.

2. When *Ss* judge alone twice, the second set of judgments is also slightly lower than the first, although here too the trend is not significant. The unidirectionality of the shift precludes the possibility of the operation of the central tendency of judgment under these conditions.

3. When the social session is also the second session, the trend for lowered judgments which is characteristic of both of these factors is reinforced so that the judgments under this condition are reliably lower than those given by the same *Ss* in the first alone situation.

4. When the social session is the first session, judgments are given which resemble second individual judgments more closely than social judgments. Where the trends for social and temporal effects operate in opposed directions, then the judgments follow the trend demanded by the temporal factor. The order of judgment is therefore more important than whether the judgments are given socially or individually.

5. There is no constriction in the range of the socially rendered judgments.

(Manuscript received March 11, 1946)

References

1. Allport, F. H. *Social psychology.* New York: Houghton-Mifflin, 1924.
2. Allport, F. H. The j-curve hypothesis of conforming behavior. *J. soc. Psychol.*, 1934, 5, 141–183.

3. CHIN, R. An analysis of conformity behavior. *Arch. Psychol.*, 1943, No. 289.
4. FARNSWORTH, P. R., & BEHNER, A. A note on the attitude of social conformity. *J. soc. Psychol.*, 1931, **2**, 126–128.
5. FEARING, F., & KRISE, E. M. Conforming behavior and the j-curve hypothesis. *J. soc. Psychol.*, 1941, **14**, 109–118.
6. GALLUP, G. Question wording in public opinion polls. *Sociometry*, 1941, **3**, 259–268.
7. PERL, R. E. The influence of a social factor upon the appreciation of humor. *Amer. J. Psychol.*, 1933, **45**, 308–312.
8. SCHONBAR, R. A. The interaction of observer-pairs in judging visual extent and movement: the formation of social norms in 'structured' situations. *Arch. Psychol.*, 1945, No. 299.
9. SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, 1935, No. 187.
10. SHERIF, M. *The psychology of social norms.* New York: Harper's, 1936.

PROACTIVE INHIBITION IN THE RECOGNITION OF NONSENSE SYLLABLES

BY HELEN E. PEIXOTTO

Wheaton College, Norton, Massachusetts

Introduction

There have been relatively few studies on inhibition as it relates to recognition. However, most of the recent experiments have shown inhibitory effects, although these effects have been slight as compared with those found when the measure of memory was recall. Most of these studies have followed the conventional experimental paradigm for retroactive inhibition; only two, those of Gibson and Raffel (**2**) and Philip (**9**) have investigated intraserial inhibition as described by Foucault (**1**). These investigators studied the effect on retention of visually perceived form as a function of its position in a series, using as their material geometric forms rather than verbal material.

The present experiment is a study of '*inhibition interne*,' to use Foucault's terminology, i.e., the study of intraserial effects during the learning of a single task, where the method of measuring memory is recognition and the material is nonsense syllables. A similar investigation was carried out by the author (**8**), where the measure of memory was reproduction. There are, however, several major differences in the procedure between the two experiments, so that the results are comparable only in part.

It might be pointed out that this paper is not a report of the experiment as originally planned, but reports data which were unexpectedly found in working out another problem. Hence, some of the details of procedure which appear to be unnecessary to the present experiment will be explained.

Procedure

The syllables used in this experiment were chosen so as to have a basis of comparison for association values. Therefore, those syllables common to the association lists of Hull (**4**), Krueger (**3**) and Glaze (**3**) were used. There were 300 syllables which appeared in all three lists. These syllables were divided randomly into 10 sets of 30 key syllables each. Five of these sets were presented to a group of college students taking introductory psychology and the remaining five sets to another similar group. The number of *S*s in each group ranged from 42 to 58, as a few *S*s missed some sessions. Each group was given one set on each of five different days, one day intervening between meetings except where the weekend occurred.

The method of presenting the syllables was to give each *S* a mimeographed booklet of four pages. The first page gave the directions for the experiment, which were read aloud by the *E* at the beginning of each session. These directions were:

> This is an experiment in learning nonsense syllables. Read the syllables by columns on pages 1, 2, and 3. Read carefully, trying to remember them. You will be given one minute to read each page. When I say 'Turn,' turn your page and read the syllables on the next page. On page 4 you will find a great many more syllables. Try to recognize those syllables which you learned on the preceding pages. Draw a line through those syllables which you recognize as having seen before. There is no time limit for page four.

On each of the first three pages, a set of 10 key syllables appeared, so varied that there was a random distribution of the various letters in the alphabet as they occurred in initial, middle or terminal places. A time limit of one min. was assigned for each of these pages in order to yield an average exposure time of two sec. per syllable, in conformity with most experiments. Since it would have been too distracting to have a time call every two sec., time was called at the end of each min. On page four 120 syllables were presented, the 30 key syllables which had just been learned and 90 others which will be called padding syllables. There was, therefore, one chance in four of accidently getting the right response. The 90 padding syllables were chosen from Glaze's list, and there were some from each list of association values. No syllable was used which had as its initial, middle or terminal letter a letter which was not used in the 30 key syllables. The 30 key syllables were distributed among the five columns, six to each column. The measure of learning was, obviously, recognition.

To summarize, then, there were 300 syllables divided into 10 sets of 30 syllables each. There were two groups of students, designated as Group A and Group B. Each group learned five sets of syllables, one set per session, and learning was measured by recognition.

The responses were checked in the following manner: a key syllable was right if it was recognized. Syllables that were falsely recognized were counted as wrong responses. Each paper was marked for the total number of right responses and the total number of wrong responses; for the number right in each column and the number wrong in each column; for the number right in each position in the five columns, that is the number of times the first key syllable in each column was right, and likewise the second and third syllables, etc.

Results

The total number of right responses and the total number of wrong responses were found for each set for both groups and these totals were averaged. The totals for the two groups were then pooled and averaged. These data are given in Table I and Fig. 1. Group B made more responses, both right and wrong, than did Group A on every trial. The form of the curves for both groups is so similar that one is justified in pooling the results. It may also be seen from the right responses in this table that in the first two sessions for Group A, and in the first three sessions for Group B, there is an improvement, after which their efficiency decreases. Group A makes least wrong responses in session four and most right responses in session two, while Group B makes least wrong responses in session three, the session in which it shows the greatest number of recognitions.

The progress of recognition from session to session may be determined in several ways: by a study either of average number of right responses, or of the average number of wrong responses, or of the average number of right responses corrected for chance by the formula $R - W/3$. These values are all given in Fig. 1. It may be

TABLE I

RIGHT AND WRONG RESPONSES FOR EACH SET

	I		II		III		IV		V	
	R	W	R	W	R	W	R	W	R	W
Ave. A	19.3	12.3	19.9	10.9	19.3	10.8	17.7	10.2	16.8	11.0
Ave. B	20.1	17.6	20.4	15.3	20.6	12.0	18.9	12.8	18.4	13.2
M	19.7	15.8	20.1	13.0	19.9	11.4	18.3	11.4	17.6	12.0
σ_M	.383	1.00	.397	.89	.407	.81	.408	.85	.466	1.01
C	14.4		15.8		16.1		14.5		13.6	
N	103		100		98		109		96	

Significance of the Differences between Sets

Right Responses: At the 1 percent level, Sets 1–5, 2–4, 2–5, 3–4, 3–5
At the 5 percent level, Sets 1–4.
Wrong Responses: At the 1 percent level, Sets 1–3, 1–4, 1–5
At the 5 percent level, Sets 1–2.

Legend: R—Right responses, W—Wrong responses, A—Group A, B—Group B, M—Mean, C—Score corrected for chance success

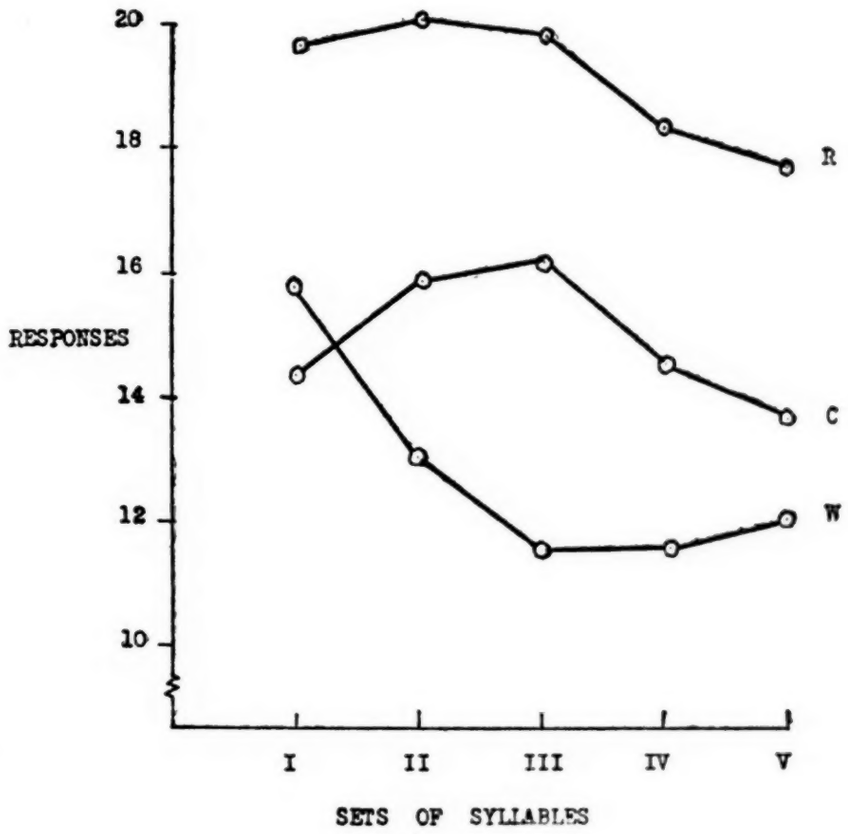

FIG. 1. Average number of right and wrong responses for each set. R—Right responses, W—Wrong responses, C—Scores corrected for chance success

seen that recognition, as determined by the number of right responses, reaches a peak in the second session and falls off considerably during the two last sessions. Recognition, as determined by the number of wrong responses, improves rapidly to the third session, then deteriorates slightly on the last session. As measured by the number of right responses corrected for chance, recognition improves to the third session, then falls off rapidly in the last two sessions.

The reliabilities of the differences found between the various sets for both right and wrong responses are given in Table I. It will be noted that differences between right responses are more significant than differences between wrong responses. These computations were not corrected for intercorrelation effects, so they may be presumed to represent the minimum significance between sets. This is also true for all the subsequent reliabilities reported.

In order to study the learning process in recognition in more detail, the average number of right and wrong responses for each column on each set was found. There are, of course, fluctuations among the sets, but on the whole the results are very similar for the right responses. The average for the five sets, therefore, expresses the relationship fairly well. The results for the wrong responses show more discrepancies, so that the average for the five sets smooths out the data, as represented in Table II and Fig. 2.

TABLE II

AVERAGE NUMBER OF RIGHT AND WRONG RESPONSES PER COLUMN

Right Responses

Column	I	II	III	IV	V	*N*
Set 1	5.11	4.23	3.70	3.82	2.97	103
Set 2	4.70	4.70	4.17	3.59	2.98	100
Set 3	4.77	4.24	3.90	3.45	3.57	98
Set 4	4.67	4.04	3.67	3.00	2.89	109
Set 5	4.32	3.69	3.74	3.15	2.76	96
Mean	4.714	4.180	3.836	3.402	3.034	507
σ_M	.048	.058	.057	.059	.061	
			Wrong Responses			
Set 1	3.34	4.37	2.86	2.43	2.79	103
Set 2	2.43	3.00	3.04	2.45	2.12	100
Set 3	2.12	2.43	2.66	2.30	2.00	98
Set 4	2.75	1.96	2.37	2.43	1.89	109
Set 5	2.82	2.61	2.75	2.03	2.07	96
Mean	2.692	2.874	2.736	2.328	2.174	507
σ_M	.099	.110	.113	.089	.092	

Significance of the Differences between Columns (for Means)

Right Responses: All differences are highly significant beyond the 1 percent level.

Wrong Responses: At the 1 percent level, Columns 1–4, 1–5, 2–4, 2–5, 3–5

At the 5 percent level, Columns 3–4

The curve for right responses is linear, with most right responses occurring in the first column, least in the last column. This is true for each set. It is interesting to note that in comparison with the other sets the first set has the greatest number of right responses for the first column, although one would not expect this from Table I. As was mentioned above, the wrong responses show more variability from column to column and among the sets than do the right responses. Here the first set has the greatest number of wrong responses, and most of these occur in the second column. The fourth set shows comparatively few wrong responses, especially in the second column and the fifth set has the least number in the fourth column.

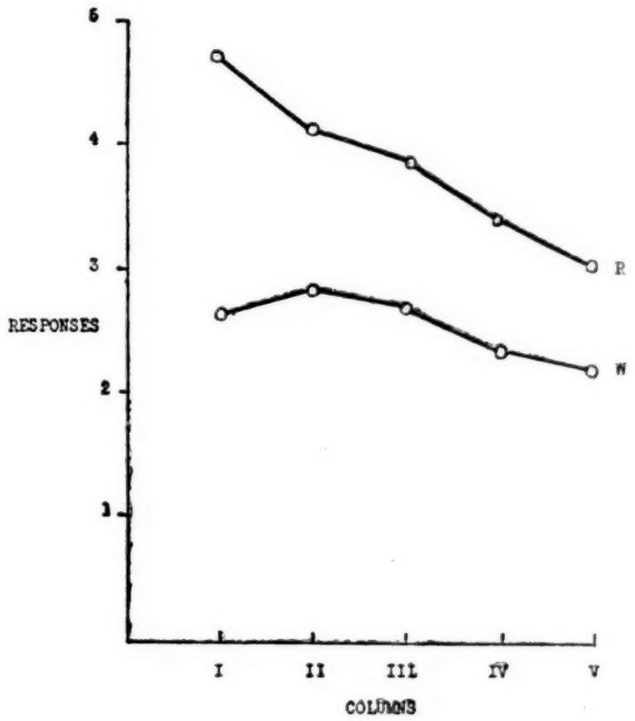

FIG. 2. Average number of right and wrong responses per column
R—Right responses, W—Wrong responses

The other columns show little variation among the sets. The average curve for these data is practically level from columns one to three, drops from columns three to four, and then levels off again. These data are highly reliable, particularly for right responses, as may be seen from Table II.

Analyzing the data further according to position of key syllables within the columns, the number of correct responses for each key syllable was found according as it was first, second, etc., within a column. Obviously there are no wrong responses, as a response to a key syllable could be only right or omitted. These data are presented in Table III and Fig. 3. Here the result is straight line, especially

TABLE III

AVERAGE NUMBER OF RESPONSES TO KEY SYLLABLES ACCORDING TO POSITION IN COLUMN

Position	1	2	3	4	5	6
Mean	34.2	33.9	31.3	31.0	28.6	30.9
N	50	50	50	50	50	50
σ_M	1.13	1.11	1.23	1.29	1.24	1.00

Significance of the Differences Between Positions
At the 1 percent level, Positions 1–5, 2–5
At the 5 percent level, Positions 1–4, 1–6, 2–6

when the results are averaged, and as they appear in Fig. 3. The first position is the highest, the succeeding positions gradually falling off through the fifth position, but the sixth position shows an increase in the number of responses. The average curve shown in Fig. 3 represents the data quite adequately, the significant differences at the one percent level occurring between the first two positions and the fifth position, as can be seen in Table III.

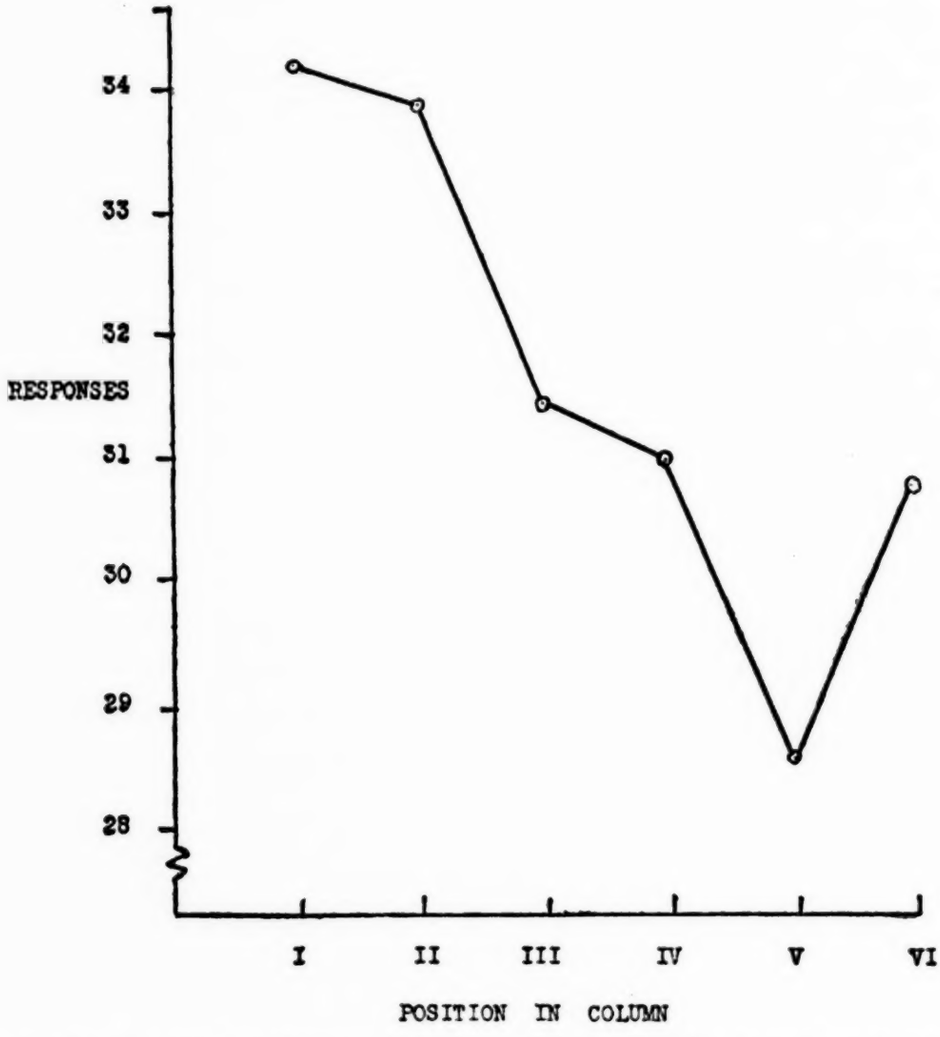

FIG. 3. Average number of responses to key syllables according to position in column

The learning process in recognition has been studied from set to set, i.e., from session to session, from column to column within the session, and from the first to sixth key syllable within the column. Using a correction for the number of right responses by the formula $R - W/3$ it is obvious that learning proceeds slowly from the first to the third session, and falls off rapidly in the last two sessions. What is remarkable is that in each session and each column, learning, as measured by the number of right responses, should fall off linearly from beginning to end. Obviously, therefore, while practice effects, in the sense of improvement in recognition, may account for some of the effects from session to session, some other, more potent factor is at work within the session and within the columns, and, as judged by the difference in slope of the learning curves, is more pronounced in the session than in the column.

Discussion

All of the data in this experiment indicate, to a greater or lesser degree, inhibitory effects on the recognition process. When the responses are considered over a period of time, i.e., from session to session, as is the case in Fig. 1, there appears to be an initial practice effect, which results in a certain amount of improvement in the first three sessions. This finding is consistent with that of other investigators, notably McGeoch (**6**). The practice effect is evident in both right and wrong responses, in the former by a gradual increase in the number of right responses, in the latter by a decrease in the number of wrong responses on the first three sessions. After maximum efficiency has been reached, proactive effects become obvious. There are fewer right responses for succeeding sessions and a tendency for more wrong responses.

An explanation of Fig. 1 in terms of practice and proactive inhibition seems most plausible, for as the experiment progresses there are not only the inhibitory effects of the syllables in the current sessions, but also cumulative inhibitory effects from those of the preceding sessions. In other words, as each set of syllables is presented to the Ss it acts as an inhibitory influence on the succeeding sets, causing interference in the recognition of the subsequent syllables. There seem to be two phenomena here, 'internal progressive inhibition' and 'external progressive inhibition,' to use the terminology of Foucault (**1**). Internal progressive inhibition, inhibition which has its origin within the original learning-memory situation, is caused chiefly by the padding syllables within each set, while external progressive inhibition, inhibition resulting from extrapolated material, is caused by all the syllables of the preceding sets. The effect of these syllables

interferes with the ability to recognize subsequent syllables, and is shown to be cumulative by the decreasing number of right responses and the increasing number of wrong responses on the final sessions. These proactive effects appear greater as the number of sessions increases, both for right responses and for wrong responses.

Apparently the methodology used in recognition experiments, and the procedure herein seems typical, is such as to enhance proactive effects, for obviously the number of interfering stimuli increases, and hence the inhibition increases, as the experiment progresses. This is true not only from session to session, but within any given session. It seems doubtful that, over a period of time during which new syllables were being presented to the *Ss* to be learned and recognized, there would be inhibitory effects on previously recognized material unless those syllables from previous sessions were included in the current session.

These proactive effects also occur within a single session, for proactive inhibition also characterizes the curves in Fig. 2, which shows the average number of right and wrong responses for each column within a set. Whatever other factors may have been operating to improve the score, they were not strong enough to influence the course of the recognition process as shown by this curve. Here, of course, the inhibition is entirely 'internal,' since the effects are the cumulative interference of the syllables in the first column on the ability to recognize those in the second, and so on to the fifth column. As a matter of fact, from studying Table II one can find some evidence for 'external' inhibition by analyzing the relative number of responses for the successive sets of syllables for each column. However, this type of inhibition is shown much more clearly in Fig. 1. There is no practice effect shown in this curve; the greatest number of right responses occur in the first column, the least in the fifth column, and the curve in Fig. 2 is linear. Proactive inhibition is, then, by far the dominant factor in the deterioration of the recognition process.

Concerning the wrong responses, however, it is evident from this analysis that practice effects are operating in the total situation, although they are not apparent in the analysis of the right responses. There are more wrong responses on the first trials than on the last. Therefore, the practice effects show themselves in greater ability in the techniques of dealing with the recognition situation, but not in the ability to recognize more syllables. In other words, by practice the *Ss* have learned to eliminate wrong responses, to make fewer false recognitions, but have not learned to make more right recognitions. The interference of the preceding material persists, and the more material the greater the interference. Thus, there is an increased amount of proactive inhibition in the recognition process in spite of

the fact that the Ss have learned to eliminate, partially at least, false-recognitions. One might interpret the curve for wrong responses in Fig. 2 as a curve of retroactive inhibition, on the grounds that, by constant elimination of responses as one proceeds from the first to the last column, there is less interference from previous material on the last column than on the first, since each response means an elimination of material which the S was trying to hold in memory, and thus each response will mean less stimuli interfering on subsequent syllables. Such an explanation is clearly in accord with the data, but the explanation in terms of practice seems to be a simpler explanation, equally in accord with the data and therefore the more acceptable. Also, an explanation in terms of practice is in accord with the results shown in Fig. 1, and it seems more probable that the same two factors should be operating throughout the experiment than that a third should be responsible for a small part of the results.

An analysis of the data according to the position of the syllables within the column shows very little variation from position to position, although the curve in Fig. 3 substantiates those in Figs. 1 and 2 by showing the same general tendency, and some of the differences are significant as indicated in Table III. Therefore, what variation there is from a straight line can also be interpreted in terms of proactive inhibition. It appears, from this curve, that within each column in a set there is a tendency for the padding syllables to influence the recognition, i.e., each column tends to be reacted to as a unit, so that the first syllable of each column is recognized with greater ease than the subsequent syllables within the particular column. In view of the tendency for this curve to rise on the last position it would be interesting to know what form the curve would take on longer lists. In this experiment this tendency does not warrant being considered more than a chance result. However, from Table III one might venture a partial explanation in terms of directional set, partly due to attentional factors of primacy and recency because of a tendency for the first and last positions to be favored. But from the data available this explanation can be only a conjecture.

As was mentioned above, the structure of the present experiment was such as to emphasize proactive effects, should inhibitory phenomenon be present at all. It is impossible to say whether retroactive inhibition was present or not, because the random presentation of the syllables in the learning situation makes it impossible to measure this form of inhibition, and it is only in the learning presentation that there is subsequent material which might act on antecedent material. In other words, to measure the retroactive inhibitory effects it would be necessary to present the material to be learned in a constant order, since, in intraserial material, it is only in the learning part of

the experiment that the consequent material acts on the antecedent material; whereas in the recall part of the experiment, especially when the measure of recall is recognition with all that this entails, it is the antecedent material which acts on the consequent material.

The results of this experiment do not seem in accord to any great degree with other similar experiments (**2, 9**). In these experiments inhibitory effects were slight, and both retroactive and proactive effects were present. Mental set or anticipation and similarity of stimuli (**9**) do not seem greatly to influence these results, but this may be because it was not possible to isolate them and not because they were not present. The discrepancies in the results between these experiments and the present experiment are probably due, at least in part, to the differences in experimental procedure. There is a certain amount of consistency between the results of the present experiment and an experiment (**8**) where inhibitory effects in the learning and recitation were isolated. The consistency occurs in the presence of practice effects in recitation in the latter experiment and in similar effects in the present experiment. It seems possible that McKinney (**7**) might have found proactive effects in the recognition process of his experiment if his data had been analyzed for intraserial effects, but there is no mention of such effects having been present or investigated.

Conclusion

The very nature of the method by which recognition is measured in memory experiments should result in proactive inhibitory effects in the recognition process. This generalization was found to be true in the present experiment, and may gain in significance since it was not the original purpose of the experiment to investigate such effects. These proactive effects are present not only within a given session, but are also cumulative over a number of sessions covering several days. They are not, however, as apparent or as effective in serial position within a session as they are in regard to the larger units measured.

Proactive effects here found seem to be most adequately explained by interference during the recognition period, which is occasioned by the padding material from among which the correct syllables are to be recognized. There is evidence that the ability to recognize syllables improves with practice. However, the improvement is not sufficiently great to counteract the inhibitory effects on recognition of right responses, but is most effective in the elimination of false recognitions. It is to be noted particularly that the practice effects, although present from session to session, are not noticeable within a session for right responses. This seems of importance since, in the

experimental procedures where memory is measured by recognition, there is usually only one session.

The results of this experiment, while not contradictory to those of similar experiments, do not closely agree with them. The discrepancies are largely in regard to the presence of retroactive effects, the difference in the strength of proactive effects, and the presence of practice effects. These discrepancies, however, may probably be explained in terms of the different experimental procedures employed.

The general conclusion, then, to be drawn from this experiment is that recognition is very definitely influenced by proactive inhibition and that the effect of practice, chiefly seen in the diminishing of the number of false recognitions, is very limited.

(Manuscript received March 4, 1946)

References

1. Foucault, M. Les inhibitions internes de fixation. *Année Psychol.*, 1928, **29**, 92–112.
2. Gibson, J. J., & Raffel, G. A technique for investigating retroactive and other inhibitory effects in immediate memory. *J. gen. Psychol.*, 1936, **15**, 107–115.
3. Glaze, J. A. The associative value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.
4. Hull, C. L. The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.*, 1933, **45**, 730–734.
5. Krueger, W. C. F. The relative difficulty of nonsense syllables. *J. exp. Psychol.*, 1934, **17**, 145–153.
6. McGeoch, J. A. The influence of associative value upon the difficulty of nonsense syllable lists. *J. genet. Psychol.*, 1930, **37**, 421–426.
7. McKinney, F. Retroactive inhibition and recognition memory. *J. exp. Psychol.*, 1935, **18**, 585–598.
8. Peixotto, H. E. Intraserial inhibition as measured by reproduction. *J. exp. Psychol.*, 1942, **31**, 17–35.
9. Philip, B. R. Proactive and retroactive effects in the recognition of form. *J. exp. Psychol.*, 1940, **26**, 502–513.

AN APPARATUS FOR THE MEASUREMENT OF CONTINUOUS CHANGES IN PALMAR SKIN RESISTANCE

BY

ERNEST A. HAGGARD

University of California

AND

RALPH GERBRANDS

Harvard University

The apparatus described in this paper was used to measure continuous changes in palmar skin resistance. Numerous methods for obtaining such records have been described in the literature, and range from the crude and innacurate to the relatively precise. Some of the more familiar methods include those of visually noting the magnitude of the deflection of a galvanometer needle, the manual tracing of a beam of light reflected from the galvanometer element onto a moving paper roll, and the photographing of such a beam of light by the use of sensitized film or paper.

An improvement on these methods, suggested to the writers by S. S. Stevens, makes use of two standard electronic instruments which are commercially available. These devices provide a simple, reliable means of obtaining a continuous graphic (pen and ink) recording of skin resistance. This apparatus, its calibration, and its working characteristics will now be described.

I. The Apparatus

(*a*) *The electrode*, as shown in Fig. 1, fits snugly against the palm of the *S*'s hand, and is held in place by a soft one-in. rubber band, and the tension necessary for a firm but convenient pressure can easily be obtained by placing the binding-post through an appropriate hole in the rubber band.[1] This electrode was obtained from M. A. Wenger.

The electrode unit shown in Fig. 1 was protected by a small ply-wood box, which was strapped to the *S*'s hand, in order to keep the electrode's position and pressure constant during the time records were being taken, but did not prevent the *S* from moving his hand about in order to be comfortable. The box was 4 in. wide, 7 in. long, and 2½ in. deep, and had an opening in the top side into which the electrode fitted, but the electrode was free from all contact with the box. The top side of the box, which was next to the *S*'s palm and fingers, should be covered with some

[1] This rubber band could be fastened most easily over the end of the binding-post, even though it is placed under the binding-post in Fig. 1. The holes in this rubber band may be made with a hand-punch used to make holes in paper which is to be put in a loose-leaf notebook.

material such as the rubber used on ping-pong paddles, or crepe rubber, to keep the *S*'s hand from moving about in relation to the box. If the electrode were to come in contact with the side of the box, there would be some change in its position or pressure, which would distort the record; a sudden movement of the hand which resulted in loss of contact between the electrode and the skin surface might burn out the galvanometer.

(*b*) *External current.*—An external current of 2.25 microamps. through the *S* was produced by a 45-volt battery connected across a circuit consisting of a 2-megohm resistor connected in series with the *S*. The large series resistor is needed to minimize variations in current caused by changes in the *S*'s resistance.

(*c*) *The recording unit.*—The recorder was a combination of a R.C.A. VoltOhmyst (Type No. 165–A) with a General Electric photoelectric recording microammeter (Cat. No. 32620; 5 microamps. full scale). These instruments were combined to form a recording d.c. vacuum-tube voltmeter with which the voltage across the *S*s electrodes could be graphically recorded and evaluated in terms of resistance.

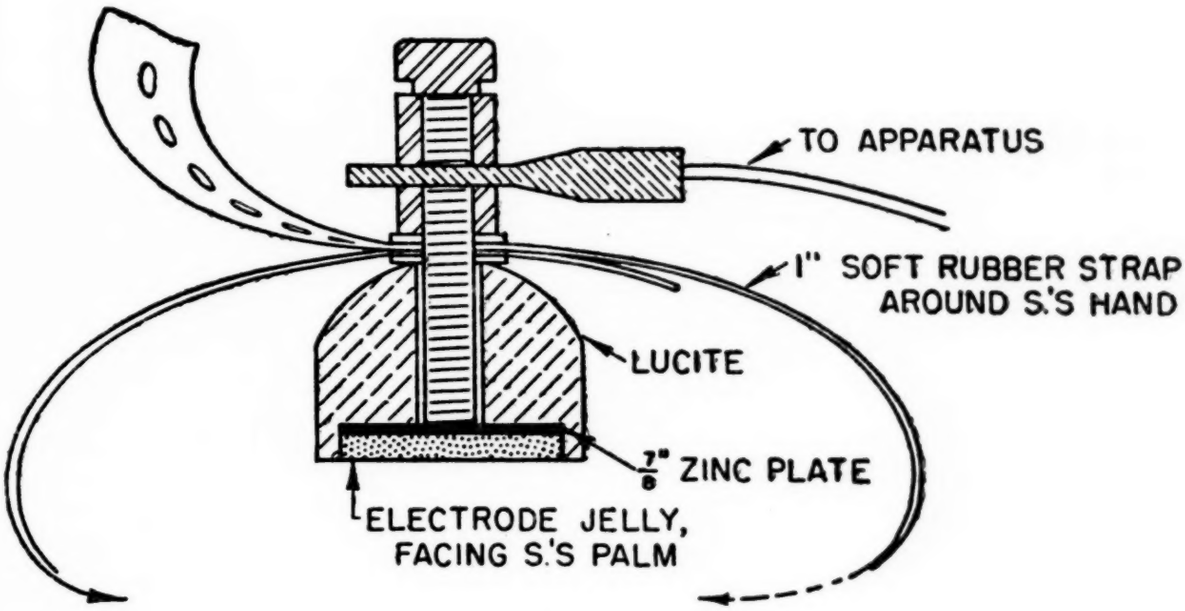

FIG. 1. Electrode used in the recording of continuous changes in palmar skin resistance. Full scale.

In order to measure both the steady component of the *S*'s voltage (i.e., general level) and the variation caused by the skin reflex (i.e., GSR), it was necessary to modify the VoltOhmyst circuit, as shown in Fig. 2. In this figure the changes made have been underlined, and the non-underlined portions of the diagram are as originally prepared by R.C.A. As shown in Fig. 2, a 10,000 ohm resistor has been added, and the 'zero adjustment' potentiometer R–21[2] has been increased to 10,000 ohms. With these changes it is possible to balance the VoltOhmyst circuit to zero, or some other arbitrary meter reading, in the presence of a steady d.c. voltage applied to the input terminals of V–1. The amount by which it is necessary to adjust the zero adjustment to obtain a balance will be a measure of the resistance across which the input voltage appears. Hence this potentiometer can be calibrated in terms of the *S*'s resistance. With the changes as shown in Fig. 2 it was possible to record resistance values up to 80,000 ohms.

Since a *S*'s initial resistance level may vary from a value as low as approximately 5000 ohms to approximately 80,000 ohms in the experiments in which this apparatus was used, it is obvious that the meter on the VoltOhmyst is too insensitive to indicate the small variations caused by the skin reflex. Therefore, in order to indicate and record these small variations the recording

[2] The potentiometer (R–21) is found on the original wiring diagram prepared by R.C.A., but does not appear in Fig. 2. This potentiometer will henceforth be referred to as the zero adjustment.

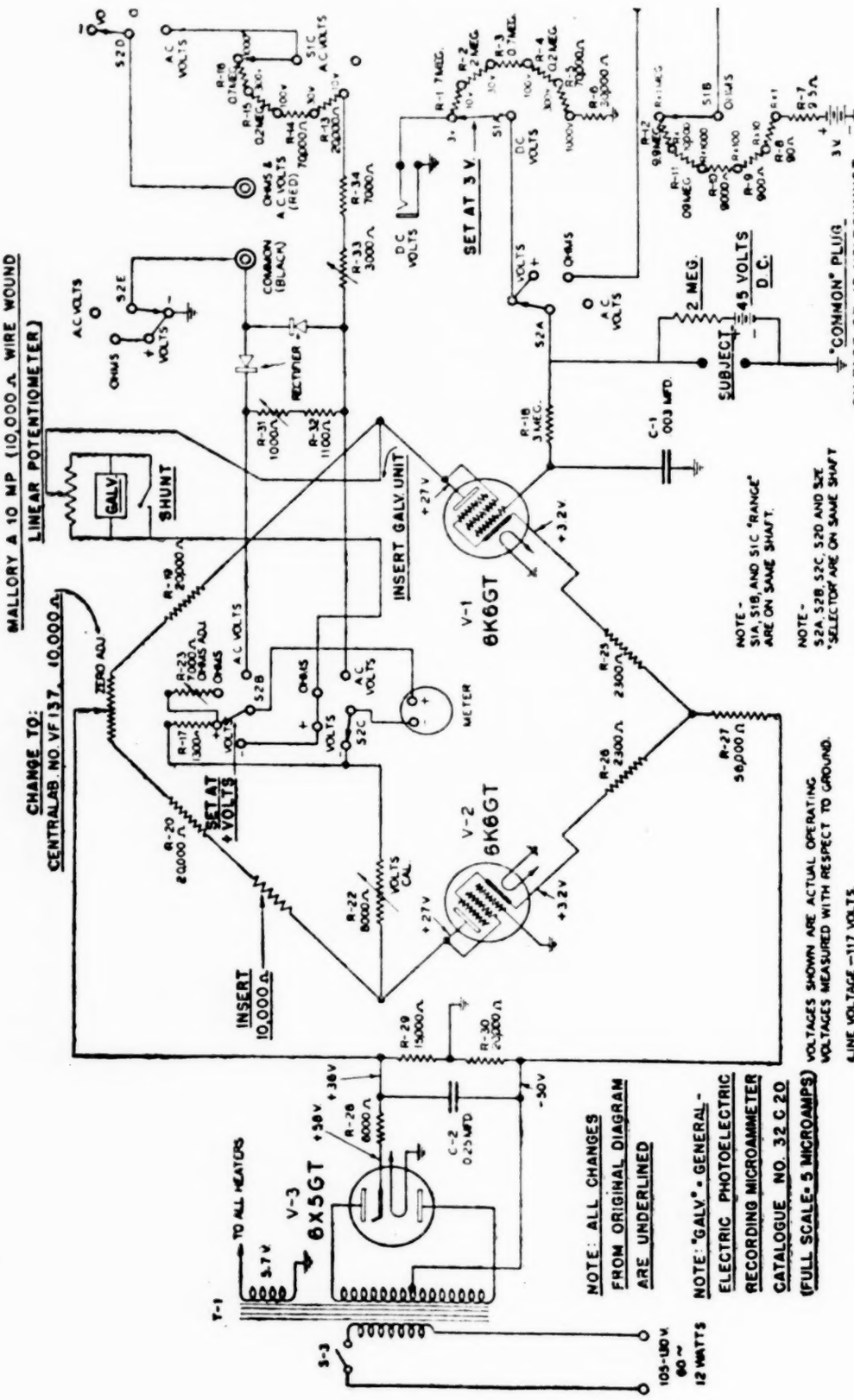
MALLORY A 10 MP (10,000Ω WIRE WOUND
LINEAR POTENTIOMETER)
CHANGE TO:
CENTRALAB NO. VF 137 10,000Ω
INSERT 10,000Ω
INSERT GALV UNIT
SET AT +VOLTS
SET AT 3 V
GALV
SHUNT
ZERO ADJ
VOLTS CAL
METER
RECTIFIER
COMMON (BLACK)
2 MEG.
45 VOLTS D.C.
SUBJECT
"COMMON" PLUG
ON FACE OF JR VOLTOHMYST
V-1 6K6GT
V-2 6K6GT
V-3 6X5GT
TO ALL HEATERS
5.7 V
T-1
S-3
105-130 V. 60 ~ 12 WATTS
C-1 .003 MFD
C-2 .025 MFD
NOTE: ALL CHANGES FROM ORIGINAL DIAGRAM ARE UNDERLINED
NOTE: "GALV" - GENERAL - ELECTRIC PHOTOELECTRIC RECORDING MICROAMMETER CATALOGUE NO. 32 C 20 (FULL SCALE = 5 MICROAMPS)
VOLTAGES SHOWN ARE ACTUAL OPERATING VOLTAGES MEASURED WITH RESPECT TO GROUND.
LINE VOLTAGE = 117 VOLTS
SELECTOR SWITCH IN "+VOLTS" POSITION
NOTE - S1A, S1B, AND S1C "RANGE" ARE ON SAME SHAFT.
NOTE - S2A, S2B, S2C, S2D AND S2E "SELECTOR" ARE ON SAME SHAFT
NOTE - R-21 & R-23 ARE ON FRONT PANEL

FIG. 2. Wiring diagram of the Junior VoltOhmyst modified to record resistances up to 80,000 ohms. Modifications include changed potentiometers, insertion of resistances, the recording unit (ink-writing G.E. photoelectric microammeter), etc. All changes in the original circuit are underlined in this figure.

microammeter was connected in series with the VoltOhmyst meter. The sensitivity of the recorder was controlled by a continuously variable shunt. This shunt was calibrated for use in two positions, high and low sensitivity. An auxiliary marking pen was attached to the G.E. recorder in order to note the presentation of the stimulus, etc.

II. Calibration of the Unit

The apparatus was calibrated with a decade resistance box connected in the place of the *S*. The calibrating procedure was carried out as follows: For *low* sensitivity the shunt of the G.E. recorder was adjusted so that for a setting of 5000 ohms on the decade resistor, a change of 2000 ohms would produce a pen deflection of one full scale division. The decade resistor was then reset to 5000 ohms and the zero adjustment of the VoltOhmyst was adjusted until the recorder pen rested at the No. 6 position of the rule chart paper (see Fig. 3). This setting of the zero adjustment was then marked 5000 ohms. The reading of the VoltOhmyst meter was then noted and taken as the reference reading for calibrating the remainder of the low sensitivity scale. That is to say, the position to which the zero adjustment had to be set in order to bring the VoltOhmyst meter to the reference value was marked on the dial of the zero adjustment for each value of input resistance used in the calibration. These values were 5, 10, 15, 20, 25, 30, 40, 50, 60, and 70 thousand ohms.

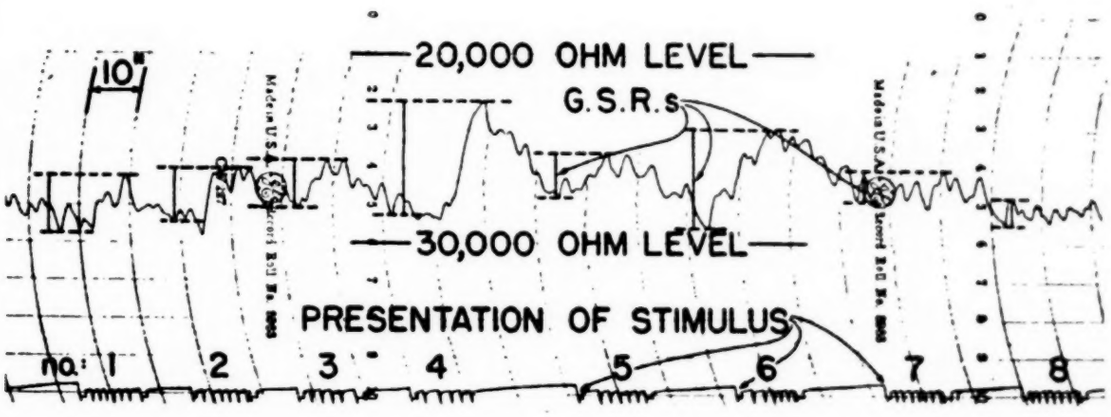

Fig. 3. Sample record of the resistance changes on the low sensitivity scale. (Figure reproduced from Haggard and Garner, 4, p. 62.) This record shows a trace which is much more irregular than most continuous records of changes in skin resistance. It appears that this record, however, is a fairly adequate representation of the resistance changes as they occurred during the experimental situation, since these minor fluctuations are not the result of under-damping, and it is very unlikely that any parasitic voltages would be large enough to produce minor fluctuations of the size shown in this figure. It appears, rather, that the high sensitivity and the short period of this recording unit registered these slight fluctuations, whereas many of the instruments in use do not record them.

Calibration in terms of both low and high sensitivity scales was necessary because of two factors present in the recording of changes in skin resistance: (*a*) there are large individual differences among *S*s in the magnitude of their average reactivity (i.e., galvanic skin response or GSR) to a constant stimulus pattern, and (*b*) the fact that the average GSR at a lower resistance level is much smaller than is the average GSR for comparable stimuli at a higher resistance level (3, 4). Thus, the *high* sensitivity scale was calibrated in the same manner as was the low sensitivity scale, with two exceptions: (*a*) the reference level was placed on the No. 8 rather than the No. 6 line, to allow for a decrease of 8000 ohms before the needle went off the record, and (*b*) each of the dark horizontal lines on the paper was made to represent 1000 ohms rather than 2000 ohms as in the case of the low sensitivity scale. Consequently, with the high sensitivity scale, the ink-writing unit could tolerate either an increase of 2000 ohms (i.e., a shift from the No. 8 to the No. 10 line) or a decrease of 8000 ohms (i.e., a shift from the No. 8 to the No. 0 line) without changing the zero adjustment on the face of the VoltOhmyst.

III. Working Characteristics of the Unit

Once the unit was assembled and calibrated, it was possible to record precisely the *S*'s resistance level at any time, so long as the maximum resistance did not exceed 80,000 ohms. The actual use of this apparatus in an experimental situation, in which the *S*'s resistance is constantly changing, would involve the following steps: (*a*) setting the VoltOhmyst meter to the reference reading by means of the zero adjustment, and (*b*) making a note of this position, such as '30-L' (i.e., 30,000 ohms, low sensitivity), as would be the case for the sample record shown in Fig. 3. If the needle of the recording unit started to go off the paper, it would be necessary to change the zero adjustment on the VoltOhmyst to the appropriate reference position, and make a notation of this change—either on the record itself or in relation to a particular stimulus. For example, with regard to Fig. 3, if the *S*'s resistance shifted from 30,000 to 22,000 ohms (i.e., from the No. 6 to the No. 2 lines for low sensitivity), the zero adjustment could be changed from the 30,000 to the 20,000 ohm position on the VoltOhmyst, and the moment this was done, the needle of the G.E. recorder would swing from the No. 2 to the No. 7 line. This adjustment can be made within one or two seconds and causes no difficulty.

With a record of the position of the zero adjustment, plus the record obtained from the stimulus marker, one can tell immediately the precise resistance of the *S* at any particular moment during the experiment. More specifically, let us return to the sample record shown in Fig. 3. Remembering that a decrease in resistance is indicated by a swing of the needle toward the top of the record, we see that in the case of stimulus No. 3, there is a shift of from approximately 28,000 ohms to 25,600 ohms, or a decrease of 2400 ohms, whereas for stimulus No. 6, the shift is roughly from 28,400 ohms to 22,400 ohms, or a decrease of 6000 ohms.[3] In other words, each of the small horizontal lines on the record indicates a change of 400 ohms on the low sensitivity scale, and one of 200 ohms on the high sensitivity scale. The direct reading of the record is not at all difficult since the width of the record is $3\frac{3}{4}$ in. (G.E. Record Roll No. 8988).

[3] These determinations are quite arbitrary, and personal taste is bound to be a factor in the reading of such records. Furthermore, the same *E* might read the record differently on a second occasion—if he were interested in a different aspect of the record, such as the resistance value occurring a constant period after the presentation of the stimulus. The more or less arbitrary rule followed in the above case may be stated as follows: the change in resistance (i.e., ohms GSR) is taken as the difference between the highest resistance value occurring a short time (say, 2–4 sec.) after the presentation of the stimulus, and the point of lowest resistance occurring within about 20 sec. after the presentation of the stimulus. The important thing, however, is that the record gives a precise indication of the *S*'s resistance, and the optimal selection of reference points for reading the record can be determined by the interests of the *E* and the requirements of the experimental situation.

Being able to determine directly and precisely both the general level of resistance (i.e., the resistance level just preceding the presentation of the stimulus) and the magnitude of the change following the presentation of the stimulus is of some importance. In the literature, these two values have been given such terms as the 'general level of resistance' in the case of the former, and the 'GSR' in the case of the latter. Both are important. Not only does a knowledge of the *S*'s general level of resistance provide information which is of interest and value in its own right (**1**, **2**), but also, an adequate quantification of the GSR value itself cannot be obtained unless one has at his disposal the *S*'s resistance level just preceding the GSR (**3**, **4**).

In conclusion, the point has been made in this paper that one can determine precisely the *S*'s resistance level at any moment during the

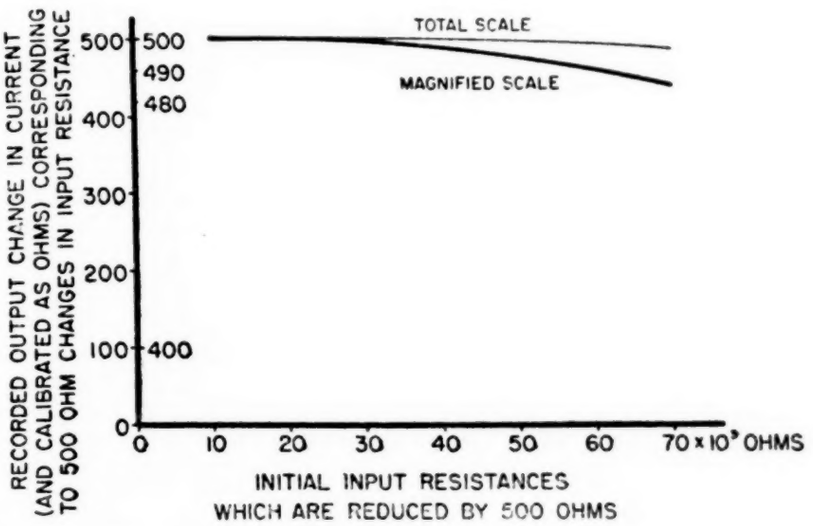

FIG. 4. Diagram showing the relation between a constant input change (a decrease of 500 ohms) and the recorded output change as registered by the recording unit over the range from 10,000 ohms to 70,000 ohms. The 'total scale' function represents the deviation from linearity on a 0–500 ohm scale (i.e., the values to the left of the ordinate), whereas the 'magnified scale' function represents the same distortion on a 400–500 ohm scale (i.e., the values to the right of the ordinate).

experiment. This statement is not quite correct. A slight distortion exists in the apparatus, and the magnitude of this distortion is shown in Fig. 4. The values for the function shown in Fig. 4 were obtained by placing a standard decade resistance box in the circuit in place of the *S*, balancing the circuit for each of the seven positions shown, and introducing a constant resistance change. More specifically, for the 10,000 and 20,000 ohm levels, the circuit was balanced and the resistance box was changed from values of 10,000 to 9500 ohms and from 20,000 to 19,500 ohms, respectively. The needle of the recording unit registered a decrease of exactly 500 ohms in each case.

However, as the resistance level was increased, an input change of 500 ohms was recorded as a change of, say, 495 ohms, and this discrepancy increased until the last value, 70,000 ohms, was reached. Here the distortion was greatest, and an input change of 500 ohms was registered as a change of 485 ohms. The 'total scale' in Fig. 4 shows this relative distortion. For all practical purposes, however, this distortion may be considered to be of negligible significance for two reasons: (*a*) for the calibrated range, the maximum distortion is only three percent for a 500 ohm change, and undoubtedly there are many uncontrolled sources of 'error' in this type of experimentation which exceed this amount, and (*b*) it has been shown that the average GSR, under constant stimulus conditions, is much greater at the higher resistance levels (**3, 4**) so that when the data have been adequately quantified, the relative distortion is only a fraction of the measured distortion shown in Fig. 4.

(Manuscript received February 6, 1946)

References

1. Haggard, E. A. Some conditions determining adjustment during and readjustment following experimentally induced stress. Ch. 40 in: *Contemporary psychopathology, a sourcebook.* (S. S. Tomkins, Ed.) Harvard Univ. Press, 1943.
2. Haggard, E. A. Experimental studies in affective processes: I. Some effects of cognitive structure and active participation on certain autonomic reactions during and following experimentally induced stress. *J. exp. Psychol.*, 1943, **33**, 257–284.
3. Haggard, E. A. Experimental studies in affective processes: II. On the quantification and evaluation of 'measured' changes in skin resistance. *J. exp. Psychol.*, 1945, **35**, 46–56.
4. Haggard, E. A., & Garner, W. R. An empirical test of a derived measure of changes in skin resistance. *J. exp. Psychol.*, 1946, **36**, 59–70.